(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2003 (12.09.2003)

PCT

(10) International Publication Number
WO 03/074723 A3

(51) International Patent Classification[7]: C12Q 1/68, C07H 21/02, 21/04

(21) International Application Number: PCT/US03/06198

(22) International Filing Date: 28 February 2003 (28.02.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/360,232	1 March 2002 (01.03.2002)	US
10/093,618	11 March 2002 (11.03.2002)	US
60/378,354	8 May 2002 (08.05.2002)	US

(71) Applicant and
(72) Inventor: **DHALLAN, Ravinder** [US/US]; 8013 Thornley Court, Bethesda, MD 20817 (US).

(74) Agents: **WITT, Eric** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
27 November 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/074723 A3

(54) Title: METHODS FOR DETECTION OF GENETIC DISORDERS

(57) Abstract: The invention provides a method useful for detection of genetic disorders. The method comprises determining the sequence of alleles of a locus of interest, and quantitating a ratio for the alleles at the locus of interest, wherein the ratio indicates the presence or absence of a chromosomal abnormality. The present invention also provides a non-invasive method for the detection of chromosomal abnormalities in a fetus. The invention is especially useful as a non-invasive method for determining the sequence of fetal DNA.

METHODS FOR DETECTION OF GENETIC DISORDERS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to U.S. Patent Application No. 10/093,618, filed March 11, 2002, and provisional U.S. Patent Application Nos. 60/360,232 and 60/378,354, filed March 1, 2002, and May 8, 2002, respectively. The contents of these applications are hereby incorporated by reference in their entirety herein.

BACKGROUND OF THE INVENTION

FIELD OF THE INVENTION

The present invention is directed to a method for the detection of genetic disorders including chromosomal abnormalities and mutations. The present invention provides a rapid, non-invasive method for determining the sequence of DNA from a fetus. The method is especially useful for detection of chromosomal abnormalities in a fetus including translocations, transversions, monosomies, trisomies, and other anueoplodies, deletions, additions, amplifications, translocations and rearrangements.

BACKGROUND ART

Chromosomal abnormalities are responsible for a significant portion of genetic defects in liveborn humans. The nucleus of a human cell contains forty-six (46) chromosomes, which contain the genetic instructions, and determine the operations of the cell. Half of the forty-six chromosomes originate from each parent. Except for the sex chromosomes, which are quite different from each other in normal males, the chromosomes from the mother and the chromosomes from the father make a matched set. The pairs were combined when the egg was fertilized by the sperm. Occasionally, an error occurs in either the formation or combination of chromosomes, and the fertilized egg is formed with too many or too few chromosomes, or with chromosomes that are mixed in some way. Because each chromosome contains many genes, chromosomal abnormalities are likely to cause serious birth defects, affecting many body systems and often including developmental disability (e.g., mental retardation).

Cells mistakenly can rejoin broken ends of chromosomes, both spontaneously and after exposure to chemical compounds, carcinogens, and irradiation. When rejoining occurs within a chromosome, a chromosome segment between the two breakpoints becomes inverted and is classified as an inversion. With inversions, there is no loss of genetic material; however, inversions can cause disruption of a critical gene, or create a fusion gene that induces a disease related condition.

In a reciprocal translocation, two non-homologous chromosomes break and exchange fragments. In this scenario, two abnormal chromosomes result: each consists of a part derived from the other chromosome and lacks a part of itself. If the translocation is of a balanced type, the individual will display no abnormal phenotypes. However, during germ-cell formation in the translocation-bearing individuals, the proper distribution of chromosomes in the egg or sperm occasionally fails, resulting in miscarriage, malformation, or mental retardation of the offspring.

In a Robertsonian translocation, the centromeres of two acrocentric (a chromosome with a non-centrally located centromere) chromosomes fuse to generate one large metacentric chromosome. The karyotype of an individual with a centric fusion has one less than the normal diploid number of chromosomes.

Errors that generate too many or too few chromosomes can also lead to disease phenotypes. For example, a missing copy of chromosome X (monosomy X) results in Turner's Syndrome, while an additional copy of chromosome 21 results in Down's Syndrome. Other diseases such as Edward's Syndrome, and Patau Syndrome are caused by an additional copy of chromosome 18, and chromosome 13, respectively.

One of the most common chromosome abnormalities is known as Down syndrome. The estimated incidence of Down's syndrome is between 1 in 1,000 to 1 in 1,100 live births. Each year approximately 3,000 to 5,000 children are born in the U.S. with this chromosomal disorder. The vast majority of children with Down syndrome (approximately 95 percent) have an extra chromosome 21. Most often, the extra chromosome originates from the mother. However, in about 3-4 percent of people with Down syndrome, a translocation between chromosome 21 and either 14 or 22 is responsible for the genetic abnormality. Finally, another chromosome problem, called mosaicism, is noted in about 1 percent of individuals with Down's syndrome. In this case, some cells have 47 chromosomes and others have 46 chromosomes. Mosiacism is thought to be the result of an error in cell division soon after conception.

Chromosomal abnormalities are congential, and therefore, prenatal diagnosis can be used to determine the health and condition of an unborn fetus. Without knowledge gained by prenatal diagnosis, there could be an untoward outcome for the fetus or the mother or both. Congenital anomalies account for 20 to 25% of perinatal deaths. Specifically, prenatal diagnosis is helpful for managing the remaining term of the pregnancy, planning for possible complications with the birth process, preparing for problems that can occur in the newborn infant, and finding conditions that may affect future pregnancies.

There are a variety of non-invasive and invasive techniques available for prenatal diagnosis including ultrasonography, amniocentesis, chorionic villus sampling (CVS), fetal blood cells in maternal blood, maternal serum alpha-fetoprotein, maternal serum beta-HCG, and maternal serum estriol. However, the techniques that are non-invasive are less specific, and the techniques with high specificity and high sensitivity are highly invasive. Furthermore, most techniques can be applied only during specific time periods during pregnancy for greatest utility.

Ultrasonography

This is a harmless, non-invasive procedure. High frequency sound waves are used to generate visible images from the pattern of the echoes made by different tissues and organs, including the fetus in the amniotic cavity. The developing embryo can be visualized at about 6 weeks of gestation. The major internal organs and extremities can be assessed to determine if any are abnormal at about 16 to 20 weeks gestation.

An ultrasound examination can be useful to determine the size and position of the fetus, the amount of amniotic fluid, and the appearance of fetal anatomy; however, there are limitations to this procedure. Subtle abnormalities, such as Down syndrome, where the morphologic abnormalities are often not marked, but only subtle, may not be detected at all.

Amniocentesis

This is a highly invasive procedure in which a needle is passed through the mother's lower abdomen into the amniotic cavity inside the uterus. This procedure can be performed at about 14 weeks gestation. For prenatal diagnosis, most amniocenteses are performed between 14 and 20 weeks gestation. However, an ultrasound examination

is performed, prior to amniocentesis, to determine gestational age, position of the fetus and placenta, and determine if enough amniotic fluid is present. Within the amniotic fluid are fetal cells (mostly derived from fetal skin) which can be grown in culture for chromosomal, biochemical, and molecular biologic analyses.

Large chromosomal abnormalities, such as extra or missing chromosomes or chromosome fragments, can be detected by karyotyping, which involves the identification and analysis of all 46 chromosomes from a cell and arranges them in their matched pairs, based on subtle differences in size and structure. In this systematic display, abnormalities in chromosome number and structure are apparent. This procedure typically takes 7-10 days for completion.

While amniocentesis can be used to provide direct genetic information, risks are associated with the procedure including fetal loss and maternal Rh sensitization. The increased risk for fetal mortality following amniocentesis is about 0.5% above what would normally be expected. Rh negative mothers can be treated with RhoGam.

Chorionic Villus Sampling (CVS)

In this procedure, a catheter is passed via the vagina through the cervix and into the uterus to the developing placenta with ultrasound guidance. The introduction of the catheter allows cells from the placental chorionic villi to be obtained and analyzed by a variety of techniques, including chromosome analysis to determine the karyotype of the fetus. The cells can also be cultured for biochemical or molecular biologic analysis. Typically, CVS is performed between 9.5 and 12.5 weeks gestation.

CVS has the disadvantage of being an invasive procedure, and it has a low but significant rate of morbidity for the fetus; this loss rate is about 0.5 to 1 % higher than for women undergoing amniocentesis. Rarely, CVS can be associated with limb defects in the fetus. Also, the possibility of maternal Rh sensitization is present. Furthermore, there is also the possibility that maternal blood cells in the developing placenta will be sampled instead of fetal cells and confound chromosome analysis.

Maternal Serum Alpha-Fetoprotein (MSAFP)

The developing fetus has two major blood proteins--albumin and alpha-fetoprotein (AFP). The mother typically has only albumin in her blood, and thus, the MSAFP test can be utilized to determine the levels of AFP from the fetus. Ordinarily,

only a small amount of AFP gains access to the amniotic fluid and crosses the placenta to mother's blood. However, if the fetus has a neural tube defect, then more AFP escapes into the amniotic fluid. Neural tube defects include anencephaly (failure of closure at the cranial end of the neural tube) and spina bifida (failure of closure at the caudal end of the neural tube). The incidence of such defects is about 1 to 2 births per 1000 in the United States. Also, if there are defects in the fetal abdominal wall, the AFP from the fetus will end up in maternal blood in higher amounts.

The amount of MSAFP increases with gestational age, and thus for the MSAFP test to provide accurate results, the gestational age must be known with certainty. Also, the race of the mother and presence of gestational diabetes can influence the level of MSAFP that is to be considered normal. The MSAFP is typically reported as multiples of the mean (MoM). The greater the MoM, the more likely a defect is present. The MSAFP test has the greatest sensitivity between 16 and 18 weeks gestation, but can be used between 15 and 22 weeks gestation. The MSAFP tends to be lower when Down's Syndrome or other chromosomal abnormalities is present.

While the MSAFP test is non-invasive, the MSAFP is not 100% specific. MSAFP can be elevated for a variety of reasons that are not related to fetal neural tube or abdominal wall defects. The most common cause for an elevated MSAFP is a wrong estimation of the gestational age of the fetus. Therefore, results from an MSAFP test are never considered definitive and conclusive.

Maternal Serum Beta-HCG

Beginning at about a week following conception and implantation of the developing embryo into the uterus, the trophoblast will produce detectable beta-HCG (the beta subunit of human chorionic gonadotropin), which can be used to diagnose pregnancy. The beta-HCG also can be quantified in maternal serum, and this can be useful early in pregnancy when threatened abortion or ectopic pregnancy is suspected, because the amount of beta-HCG will be lower than normal.

In the middle to late second trimester, the beta-HCG can be used in conjunction with the MSAFP to screen for chromosomal abnormalities, in particular for Down syndrome. An elevated beta-HCG coupled with a decreased MSAFP suggests Down syndrome. High levels of HCG suggest trophoblastic disease (molar pregnancy). The

absence of a fetus on ultrasonography along with an elevated HCG suggests a hydatidiform mole.

Maternal Serum Estriol

The amount of estriol in maternal serum is dependent upon a viable fetus, a properly functioning placenta, and maternal well-being. Dehydroepiandrosterone (DHEA) is made by the fetal adrenal glands, and is metabolized in the placenta to estriol. The estriol enters the maternal circulation and is excreted by the maternal kidney in urine or by the maternal liver in the bile. Normal levels of estriol, measured in the third trimester, will give an indication of general well-being of the fetus. If the estriol level drops, then the fetus is threatened and an immediate delivery may be necessary. Estriol tends to be lower when Down syndrome is present and when there is adrenal hypoplasia with anencephaly.

The Triple Screen Test

The triple screen test comprises analysis of maternal serum alpha-feto-protein (MSAFP), human chorionic gonadotrophin (hCG), and unconjugated estriol (uE3). The blood test is usually performed 16-18 weeks after the last menstrual period. While the triple screen test is non-invasive, abnormal test results are not indicative of a birth defect. Rather, the test only indicates an increased risk and suggests that further testing is needed. For example, 100 out of 1,000 women will have an abnormal result from the triple screen test. However, only 2-3 of the 100 women will have a fetus with a birth defect. This high incidence of false positives causes tremendous stress and unnecessary anxiety to the expectant mother.

Fetal Cells Isolated From Maternal Blood

The presence of fetal nucleated cells in maternal blood makes it possible to use these cells for noninvasive prenatal diagnosis (Walknowska, et al., Lancet 1:1119-1122, 1969; Lo et al., Lancet 2:1363-65, 1989; Lo et al., Blood 88:4390-95, 1996). The fetal cells can be sorted and analyzed by a variety of techniques to look for particular DNA sequences (Bianchi et al., Am. J. Hum. Genet. 61:822-29, (1997); Bianchi et al., PNAS 93:705-08, (1996)). Fluorescence in-situ hybridization (FISH) is one technique that can be applied to identify particular chromosomes of the fetal cells recovered from maternal

blood and diagnose aneuploid conditions such as trisomies and monosomy X. Also, it has been reported that the number of fetal cells in maternal blood increases in aneuploid pregnancies.

The method of FISH uses DNA probes labeled with colored fluorescent tags that allow detection of specific chromosomes or genes under a microscope. Using FISH, subtle genetic abnormalities that cannot be detected by standard karyotyping are readily identifiable. This procedure typically takes 24-48 hours to complete. Additionally, using a panel of multi-colored DNA FISH probes, abnormal chromosome copy numbers can be seen.

While improvements have been made for the isolation and enrichment of fetal cells, it is still difficult to get many fetal blood cells. There may not be enough to reliably determine anomalies of the fetal karyotype or assay for other abnormalities. Furthermore, most techniques are time consuming, require high-inputs of labor, and are difficult to implement for a high throughput fashion.

Fetal DNA From Maternal Blood

Fetal DNA has been detected and quantitated in maternal plasma and serum (Lo et al., Lancet 350:485-487 (1997); Lo et al., Am. J. hum. Genet. 62:768-775 (1998)). Multiple fetal cell types occur in the maternal circulation, including fetal granulocytes, lymphocytes, nucleated red blood cells, and trophoblast cells (Pertl, and Bianchi, Obstetrics and Gynecology 98: 483-490 (2001)). Fetal DNA can be detected in the serum at the seventh week of gestation, and increases with the term of the pregnancy. The fetal DNA present in the maternal serum and plasma is comparable to the concentration of DNA obtained from fetal cell isolation protocols.

Circulating fetal DNA has been used to determine the sex of the fetus (Lo et al., Am. J. hum. Genet. 62:768-775 (1998)). Also, fetal rhesus D genotype has been detected using fetal DNA. However, the diagnostic and clinical applications of circulating fetal DNA is limited to genes that are present in the fetus but not in the mother (Pertl and Bianchi, Obstetrics and Gynecology 98: 483-490 (2001)). Thus, a need still exists for a non-invasive method that can determine the sequence of fetal DNA and provide definitive diagnosis of chromosomal abnormalities in a fetus.

BRIEF SUMMARY OF THE INVENTION

The invention is directed to a method for detection of genetic disorders including mutations and chromosomal abnormalities. In a preferred embodiment, the present invention is used to detect mutations, and chromosomal abnormalities including but not limited to translocation, transversion, monosomy, trisomy, and other anueoplodies, deletion, addition, amplification, fragment, translocation, and rearrangement. Numerous abnormalities can be detected simultaneously. The present invention also provides a non-invasive method to determine the sequence of fetal DNA from a sample of a pregnant female. The present invention can be used to detect any alternation in gene sequence as compared to the wild type sequence including but not limited to point mutation, reading frame shift, transition, transversion, addition, insertion, deletion, addition-deletion, frame-shift, missense, reverse mutation, and microsatellite alteration.

In one embodiment, the present invention is directed to a method for detecting chromosomal abnormalities said method comprising: (a) determining the sequence of alleles of a locus of interest on template DNA, and (b) quantitating a ratio for the alleles at a heterozygous locus of interest that was identified from the locus of interest of (a), wherein said ratio indicates the presence or absence of a chromosomal abnormality.

In another embodiment, the present invention provides a non-invasive method for determining the sequence of a locus of interest on fetal DNA, said method comprising: (a) obtaining a sample from a pregnant female; (b) adding a cell lysis inhibitor to the sample of (a); (c) obtaining template DNA from the sample of (b), wherein said template DNA comprises fetal DNA and maternal DNA; and (d) determining the sequence of a locus of interest on template DNA.

In another embodiment, the template DNA is obtained from a sample including but not limited to a cell, tissue, blood, serum, plasma, saliva, urine, tears, vaginal secretion, umbilical cord blood, chorionic villi, amniotic fluid, embryonic tissue, embryo, a two-celled embryo, a four-celled embryo, an eight-celled embryo, a 16-celled embryo, a 32- celled embryo, a 64-celled embryo, a 128-celled embryo, a 256-celled embryo, a 512-celled embryo, a 1024-celled embryo, lymph fluid, cerebrospinal fluid, mucosa secretion, peritoneal fluid, ascitic fluid, fecal matter, or body exudate.

In one embodiment, the template DNA is obtained from a sample from a pregnant female. In a preferred embodiment, the template DNA is obtained from a pregnant human female.

In another embodiment, the template DNA is obtained from an embryo. In a preferred embodiment, the template DNA is obtained from a single cell from an embryo.

In another embodiment, a cell lysis inhibitor is added to the sample including but not limited to formaldehyde, and derivatives of formaldehyde, formalin, glutaraldehyde, and derivatives of glutaraldehyde, crosslinkers, primary amine reactive crosslinkers, sulfhydryl reactive crosslinkers, sulfhydryl addition or disulfide reduction, carbohydrate reactive crosslinkers, carboxyl reactive crosslinkers, photoreactive crosslinkers, cleavable crosslinkers, AEDP, APG, BASED, $BM(PEO)_3$, $BM(PEO)_4$, BMB, BMDB, BMH, BMOE, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP, DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, sulfo-BSOCOES, Sulfo-DST, or Sulfo-EGS.

In another embodiment, an agent that prevents DNA destruction is added to the sample including but not limited to DNase inhibitors, zinc chloride, ethylenediaminetetraacetic acid, guanidine-HCl, guanidine isothiocyanate, N-lauroylsarcosine, and Na-dodecylsulphate.

In a preferred embodiment, template DNA is obtained from the plasma of the blood from a pregnant female. In another embodiment, the template DNA is obtained from the serum of the blood from a pregnant female.

In another embodiment, template DNA comprises fetal DNA and maternal DNA.

In another embodiment, the locus of interest on the template DNA is selected from a maternal homozygous locus of interest. In another embodiment, the locus of interest on the template DNA is selected from a maternal heterozygous locus of interest.

In another embodiment, the locus of interest on the template DNA is selected from a paternal homozygous locus of interest. In another embodiment, the locus of interest on the template DNA is selected from a paternal heterozygous locus of interest.

In one embodiment, the sequence of alleles of multiple loci of interest on a single chromosome is determined. In a preferred embodiment, the sequence of alleles of multiple loci of interest on multiple chromosomes is determined.

In another embodiment, determining the sequence of alleles of a locus of interest comprises a method including but not limited to allele specific PCR, gel electrophoresis, ELISA, mass spectrometry, hybridization, primer extension, fluorescence polarization, fluorescence detection, fluorescence resonance energy transfer (FRET), sequencing, DNA microarray, southern blot, slot blot, dot blot, and MALDI-TOF mass spectrometry.

In a preferred embodiment, determining the sequence of alleles of a locus of interest comprises (a) amplifying the locus of interest using a first and second primers, wherein the second primer contains a recognition site for a restriction enzyme that generates a 5' overhang containing the locus of interest; (b) digesting the amplified DNA with the restriction enzyme that recognizes the recognition site on the second primer; (c) incorporating a nucleotide into the digested DNA of (b) by using the 5' overhang containing the locus of interest as a template; and (d) determining the sequence of the locus of interest by determining the sequence of the DNA of (c).

In one embodiment, the amplification can comprise polymerase chain reaction (PCR). In a further embodiment, the annealing temperature for cycle 1 of PCR can be about the melting temperature of the annealing length of the second primer. In another embodiment, the annealing temperature for cycle 2 of PCR can be about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. In another embodiment, the annealing temperature for the remaining cycles can be about the melting temperature of the entire sequence of the second primer.

In another embodiment, the recognition site on the second primer is for a restriction enzyme that cuts at a distance from its binding site and generates a 5' overhang, which contains the locus of interest. In a preferred embodiment, the recognition site on the second primer is for a Type IIS restriction enzyme. The Type IIS restriction enzyme includes but is not limited to Alw I, Alw26 I, Bbs I, Bbv I, BceA I, Bmr I, Bsa I, Bst71 I, BsmA I, BsmB I, BsmF I, BspM I, Ear I, Fau I, Fok I, Hga I, Ple I, Sap I, SSfaN I, and Sthi32 I, and more preferably BceA I and BsmF I.

In one embodiment, the 3' end of the second primer is adjacent to the locus of interest.

In another embodiment, the annealing length of the second primer is selected from the group consisting of 35-30, 30-25, 25-20, 20-15, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, and less than 4 bases.

In another embodiment, amplifying the loci of interest comprises using first and second primers that contain a portion of a restriction enzyme recognition site, wherein said recognition site contains at least one variable nucleotide, and after amplification the full restriction enzyme recognition site is generated, and the 3' region of said primers can contain mismatches with the template DNA, and digestion with said restriction enzyme generates a 5' overhang containing the locus of interest.

In a preferred embodiment, the recognition site for restriction enzymes including but not limited to BsaJ I (5' C$^{\downarrow}$CNNGG 3'), BssK I (5'$^{\downarrow}$CCNGG 3'), Dde I (5'C$^{\downarrow}$TNAG 3'), EcoN I (5'CCTNN$^{\downarrow}$NNNAGG 3'), Fnu4H I (5'GC$^{\downarrow}$NGC 3'), Hinf I (5'G$^{\downarrow}$ANTC 3'), PflF 1(5' GACN$^{\downarrow}$NNGTC 3'), Sau96 I (5' G$^{\downarrow}$GNCC 3'), ScrF I (5' CC$^{\downarrow}$NGG 3'), Tthl 11 I (5' GACN$^{\downarrow}$NNGTC 3'), and more preferably Fnu4H I and EcoN I, is generated after amplification.

In another embodiment, the 5' region of the first and/or second primer contains a recognition site for a restriction enzyme. In a preferred embodiment, the restriction enzyme recognition site is different from the restriction enzyme recognition site that generates a 5' overhang containing the locus of interest.

In a further embodiment, the method of the invention further comprises digesting the DNA with a restriction enzyme that recognizes the recognition site at the 5' region of the first and/or second primer.

The first and/or second primer can contain a tag at the 5' terminus. Preferably, the first primer contains a tag at the 5' terminus. The tag can be used to separate the amplified DNA from the template DNA. The tag can be used to separate the amplified DNA containing the labeled nucleotide from the amplified DNA that does not contain the labeled nucleotide. The tag, e.g., is selected from the group consisting of: radioisotope, fluorescent reporter molecule, chemiluminescent reporter molecule, antibody, antibody fragment, hapten, biotin, derivative of biotin, photobiotin, iminobiotin, digoxigenin, avidin, enzyme, acridinium, sugar, enzyme, apoenzyme, homopolymeric oligonucleotide, hormone, ferromagnetic moiety, paramagnetic moiety, diamagnetic moiety, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, moiety having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity, and combinations thereof. Preferably, the tag is biotin. The biotin tag is used to separate amplified DNA from the template DNA using a streptavidin matrix. The streptavidin matrix is coated on wells of a microtiter plate.

The incorporation of a nucleotide in the method of the invention is by a DNA polymerase including but not limited to E. coli DNA polymerase, Klenow fragment of E. coli DNA polymerase I, T7 DNA polymerase, T4 DNA polymerase, T5 DNA polymerase, Klenow class polymerases, Taq polymerase, Pfu DNA polymerase, Vent

polymerase, bacteriophage 29, REDTaq™ Genomic DNA polymerase, or sequenase. The incorporation of a nucleotide can further comprise using a mixture of labeled and unlabeled nucleotides. One nucleotide, two nucleotides, three nucleotides, four nucleotides, five nucleotides, or more than five nucleotides can be incorporated. A combination of labeled and unlabeled nucleotides can be incorporated. The labeled nucleotide is selected from the group consisting of a dideoxynucleotide triphosphate and deoxynucleotide triphosphate. The unlabeled nucleotide is selected from the group consisting of a dideoxynucleotide triphosphate and deoxynucleotide triphosphate. The labeled nucleotide is labeled with a molecule selected from the group consisting of radioactive molecule, fluorescent molecule, antibody, antibody fragment, hapten, carbohydrate, biotin, and derivative of biotin, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, and moiety, having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. Preferably, the labeled nucleotide is labeled with a fluorescent molecule. The incorporation of a fluorescent labeled nucleotide further comprises using a mixture of fluorescent and unlabeled nucleotides.

In one embodiment, the determination of the sequence of the locus of interest comprises detecting the incorporated nucleotide. In one embodiment, the detection is by a method selected from the group consisting of gel electrophoresis, capillary electrophoresis, microchannel electrophoresis, polyacrylamide gel electrophoresis, fluorescence detection, fluorescence polarization, DNA sequencing, Sanger dideoxy sequencing, ELISA, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry, fluorometry, infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry, DNA hybridization, DNA microarray, southern blot, slot blot, and dot blot.

In one embodiment, the sequence of alleles of one to tens to hundreds to thousands of loci of interest on a single chromosome on template DNA is determined. In a preferred embodiment, the sequence of alleles of one to tens to hundreds to thousands of loci of interest on multiple chromosomes is determined.

In a preferred embodiment, the locus of interest is suspected of containing a single nucleotide polymorphism or mutation. The method can be used for determining sequences of multiple loci of interest concurrently. The template DNA can comprise multiple loci from a single chromosome. The template DNA can comprise multiple loci

from different chromosomes. The loci of interest on template DNA can be amplified in one reaction. Alternatively, each of the loci of interest on template DNA can be amplified in a separate reaction. The amplified DNA can be pooled together prior to digestion of the amplified DNA. Each of the labeled DNA containing a locus of interest can be separated prior to determining the sequence of the locus of interest. In one embodiment, at least one of the loci of interest is suspected of containing a single nucleotide polymorphism or a mutation.

In another embodiment, the ratio of alleles at a heterozygous locus of interest on a chromosome is compared to the ratio of alleles at a heterozygous locus of interest on a different chromosome. There is no limitation as to the chromosomes that can be compared. The ratio for the alleles at a heterozygous locus of interest on any chromosome can be compared to the ratio for the alleles at a heterozygous locus of interest on any other chromosome. In a preferred embodiment, the ratio of alleles at multiple heterozygous loci of interest on a chromosome are summed and compared to the ratio of alleles at multiple heterozygous loci of interest on a different chromosome.

In another embodiment, the ratio of alleles at a heterozygous locus of interest on a chromosome is compared to the ratio of alleles at a heterozygous locus of interest on two, three, four or more than four chromosomes. In another embodiment, the ratio of alleles at multiple loci of interest on a chromosome is compared to the ratio of alleles at multiple loci of interest on two, three, four, or more than four chromosomes.

In another embodiment, the ratio of the alleles at a locus of interest on a chromosome is compared to the ratio of the alleles at a locus of interest on a different chromosome, wherein a difference in the ratios indicates the presence or absence of a chromosomal abnormality. In another embodiment, the ratio of the alleles at multiple loci of interest on a chromosome is compared to the ratio of the alleles at multiple loci of interest on a different chromosome, wherein a difference in the ratios indicates the presence or absence of a chromosomal abnormality.

In another embodiment, the sequence of one to tens to hundreds to thousands of loci of interest on the template DNA obtained from a sample of a pregnant female is determined. In one embodiment, the loci of interest are on one chromosome. In another embodiment, the loci of interest are on multiple chromosomes.

BRIEF DESCRIPTION OF THE FIGURES

FIG. 1A. A Schematic diagram depicting a double stranded DNA molecule. A pair of primers, depicted as bent arrows, flank the locus of interest, depicted as a triangle symbol at base N14. The locus of interest can be a single nucleotide polymorphism, point mutation, insertion, deletion, translocation, etc. Each primer contains a restriction enzyme recognition site about 10 by from the 5' terminus depicted as region "a" in the first primer and as region "d" in the second primer. Restriction recognition site "a" can be for any type of restriction enzyme but recognition site "d" is for a restriction enzyme, which cuts "n" nucleotides away from its recognition site and leaves a 5' overhang and a recessed 3' end. Examples of such enzymes include but are not limited to BceAI and BsmF I. The 5' overhang serves as a template for incorporation of a nucleotide into the 3' recessed end.

The first primer is shown modified with biotin at the 5' end to aid in purification. The sequence of the 3' end of the primers is such that the primers anneal at a desired distance upstream and downstream of the locus of interest. The second primer anneals close to the locus of interest; the annealing site, which is depicted as region "c," is designed such that the 3' end of the second primer anneals one base away from the locus of interest. The second primer can anneal any distance from the locus of interest provided that digestion with the restriction enzyme, which recognizes the region "d" on this primer, generates a 5' overhang that contains the locus of interest. The first primer annealing site, which is depicted as region "b," is about 20 bases.

FIG. 1B. A schematic diagram depicting the annealing and extension steps of the first cycle of amplification by PCR. The first cycle of amplification is performed at about the melting temperature of the 3' region, which anneals to the template DNA, of the second primer, depicted as region "c," and is 13 base pairs in this example. At this temperature, both the first and second primers anneal to their respective complementary strands and begin extension, depicted by dotted lines. In this first cycle, the second primer extends and copies the region b where the first primer can anneal in the next cycle.

FIG. 1C. A schematic diagram depicting the annealing and extension steps following denaturation in the second cycle of amplification of PCR. The second cycle of amplification is performed at a higher annealing temperature (TM2), which is about the

melting temperature of the 20 by of the 3' region of the first primer that anneals to the template DNA, depicted as region "b." Therefore at TM2, the first primer, which contains region b' which is complementary to region b, can bind to the DNA that was copied in the first cycle of the reaction. However, at TM2 the second primer cannot anneal to the original template DNA or to DNA that was copied in the first cycle of the reaction because the annealing temperature is too high. The second primer can anneal to 13 bases in the original template DNA but TM2 is calculated at about the melting temperature of 20 bases.

FIG. 1D. A schematic diagram depicting the annealing and extension reactions after denaturation during the third cycle of amplification. In this cycle, the annealing temperature, TM3, is about the melting temperature of the entire second primer, including regions "c" and "d." The length of regions "c" + "d" is about 27-33 by long, and thus TM3 is significantly higher than TM1 and TM2. At this higher TM the second primer, which contain regions c' and d', anneals to the copied DNA generated in cycle 2.

FIG. 1E. A schematic diagram depicting the annealing and extension reactions for the remaining cycles of amplification. The annealing temperature for the remaining cycles is TM3, which is about the melting temperature of the entire second primer. At TM3, the second primer binds to templates that contain regions c' and d' and the first primer binds to templates that contain regions a' and b. By raising the annealing temperature successively in each cycle for the first three cycles, from TM1, TM2, and TM3, nonspecific amplification is significantly reduced.

FIG. 1F. A schematic diagram depicting the amplified locus of interest bound to a solid matrix.

FIG. 1G. A schematic diagram depicting the bound, amplified DNA after digestion with restriction enzyme "d." The "downstream" end is released into the supernatant, and can be removed by washing with any suitable buffer. The upstream end containing the locus of interest remains bound to the solid matrix.

FIG. 1H. A schematic diagram depicting the bound amplified DNA, after "filling in" with a labeled ddNTP. A DNA polymerase is used to "fill in" the base (N'14) that is complementary to the locus of interest (N14). In this example, only ddNTPs are present in this reaction, such that only the locus of interest or SNP of interest is filled in.

FIG. 1I. A schematic diagram depicting the labeled, bound DNA after digestion with restriction enzyme "a." The labeled DNA is released into the supernatant, which can be collected to identify the base that was incorporated.

FIG. 2. A schematic diagram depicting double stranded DNA templates with n number of loci of interest and n number of primer pairs, x_1, y_1 to x_n, y_n, specifically annealed such that a primer flanks each locus of interest. The first primers are biotinylated at the 5' end, depicted by •, and contain a restriction enzyme recognition site, "a", which can be any type of restriction enzyme. The second primers contain a restriction enzyme recognition site, "d," where "d" is a recognition site for a restriction enzyme that cuts "n" nucleotides away from its recognition site, and generates a 5' overhang containing the locus of interest and a recessed 3' end. The second primers anneal adjacent to the respective loci of interest. The exact position of the restriction enzyme site "d" in the second primers is designed such that digesting the PCR product of each locus of interest with restriction enzyme "d" generates a 5' overhang containing the locus of interest and a 3' recessed end. The annealing sites of the first primers are about 20 bases long and are selected such that each successive first primer is further away from its respective second primer. For example, if at locus 1 the 3' ends of the first and second primers are Z base pairs apart, then at locus 2, the 3' ends of the first and second primers are Z + K base pairs apart, where K = 1, 2, 3 or more than three bases. Primers for locus N are Z_{N-1} + K base pairs apart. The purpose of making each successive first primer further apart from their respective second primers is such that the "filled in" restriction fragments (generated after amplification, purification, digestion and labeling as described in FIGS. 1B-1I) differ in size and can be resolved, for example by electrophoresis, to allow detection of each individual locus of interest.

FIG. 3. PCR amplification of SNPs using multiple annealing temperatures. A sample containing genomic DNA templates from thirty-six human volunteers was

analyzed for the following four SNPs: SNP HC21S00340 (lane 1), identification number as assigned in the Human Chromosome 21 cSNP Database, located on chromosome 21; SNP TSC 0095512 (lane 2), located on chromosome 1, SNP TSC 0214366 (lane 3), located on chromosome 1; and SNP TSC 0087315 (lane 4), located on chromosome 1. Each SNP was amplified by PCR using three different annealing temperature protocols, herein referred to as the low stringency annealing temperature; medium stringency annealing temperature; and high stringency annealing temperature. Regardless of the annealing temperature protocol, each SNP was amplified for 40 cycles of PCR. The denaturation step for each PCR reaction was performed for 30 seconds at 95°C.

FIG. 3A. Photograph of a gel demonstrating PCR amplification of the 4 different SNPs using the low stringency annealing temperature protocol.

FIG. 3B. Photograph of a gel demonstrating PCR amplification of the 4 different SNPs using medium stringency annealing temperature protocol.

FIG. 3C. Photograph of a gel demonstrating PCR amplification of the 4 different SNPs using the high stringency annealing temperature protocol.

FIG. 4A. A depiction of the DNA sequence of SNP HC21S00027, as assigned by the Human Chromosome 21 cSNP database, located on chromosome 21. A first primer and a second primer are indicated above and below, respectively, the sequence of HC21S00027. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BsmF I and contains 13 bases that anneal to the DNA sequence. The SNP is indicated by R (A/G) and r (T/C) (complementary to R).

FIG. 4B. A depiction of the DNA sequence of SNP HC21S00027, as assigned by the Human Chromosome 21 cSNP database, located on chromosome 21. A first primer and a second primer are indicated above and below, respectively, the sequence of HC21S00027. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme

recognition site for BceA I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by R (A/G) and r (T/C) (complementary to R).

FIG. 4C. A depiction of the DNA sequence of SNP TSC0095512 from chromosome 1. The first primer and the second primer are indicated above and below, respectively, the sequence of TSC0095512. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BsmF I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by S (G/C) and s (C/G) (complementary to S).

FIG. 4D. A depiction of the DNA sequence of SNP TSC0095512 from chromosome 1. The first primer and the second primer are indicated above and below, respectively, the sequence of TSC0095512. The first primer is biotinylated and contains the restriction enzyme recognition site for EcoRI. The second primer contains the restriction enzyme recognition site for BceA I and has 13 bases that anneal to the DNA sequence. The SNP is indicated by S (G/C) and s (C/G) (complementary to S).

FIGS. 5A-5D. A schematic diagram depicting the nucleotide sequences of SNP HC21S00027 (FIGS. 5A and 5B) and SNP TSC0095512 (FIGS. 5C and 5D) after amplification with the primers described in FIGS. 4A-4D. Restriction sites in the primer sequence are indicated in bold.

FIGS. 6A-6D. A schematic diagram depicting the nucleotide sequences of each amplified SNP after digestion with the appropriate Type IIS restriction enzyme. FIGS. 6A and 6B depict fragments of SNP HC21S00027 digested with the Type IIS restriction enzymes BsmF I and BceA I, respectively. FIGS. 6C and 6D depict fragments of SNP TSC0095512 digested with the Type IIS restriction enzymes BsmF I and BceA I, respectively.

FIGS. 7A-7D. A schematic diagram depicting the incorporation of a fluorescently labeled nucleotide using the 5' overhang of the digested SNP site as a template to "fill in" the 3' recessed end. FIGS. 7A and 7B depict the digested SNP HC21S00027 locus with an incorporated labeled ddNTP ($*R^{-dd}$ = fluorescent dideoxy nucleotide). FIGS. 7C and 7D depict the digested SNP TSC0095512 locus with an

incorporated labeled ddNTP (*S^{-dd} = fluorescent dideoxy nucleotide). The use of ddNTPs ensures that the 3' recessed end is extended by one nucleotide, which is complementary to the nucleotide of interest or SNP site present in the 5' overhang.

FIG. 7E. A schematic diagram depicting the incorporation of dNTPs and a ddNTP into the 5' overhang containing the SNP site. SNP HC21S00007 was digested with BsmF I, which generates a four base 5' overhang. The use of a mixture of dNTPs and ddNTPs allows the 3' recessed end to be extended one nucleotide (a ddNTP is incorporated first); two nucleotides (a dNTP is incorporated followed by a ddNTP); three nucleotides (two dNTPs are incorporated, followed by a ddNTP); or four nucleotides (three dNTPs are incorporated, followed by a ddNTP). All four products can be separated by size, and the incorporated nucleotide detected (*R^{-dd} = fluorescent dideoxy nucleotide). Detection of the first nucleotide, which corresponds to the SNP or locus site, and the next three nucleotides provides an additional level of quality assurance. The SNP is indicated by R (A/G) and r (T/C) (complementary to R).

FIGS. 8A-8D. Release of the "filled in" SNP from the solid support matrix, i.e. streptavidin coated well. SNP HC21S00027 is. shown in FIGS. 8A and 8B, while SNP TSC0095512 is shown in FIGS. 8C and 8D. The "filled in" SNP is free in solution, and can be detected.

FIG. 9A. Sequence analysis of SNP HC21S00027 digested with BceAI. Four "fill in" reactions are shown; each reaction contained one fluorescently labeled nucleotide, ddGTP, ddATP, ddTTP, or ddCTP, and unlabeled ddNTPs. The 5' overhang generated by digestion with BceA I and the expected nucleotides at this SNP site are indicated.

FIG. 9B. Sequence analysis of SNP TSC0095512. SNP TSC0095512 was amplified with a second primer that contained the recognition site for BceA I, and in a separate reaction, with a second primer that contained the recognition site for BsmF I. Four fill in reactions are shown for each PCR product; each reaction contained one fluorescently labeled nucleotide, ddGTP, ddATP, ddTTP, or ddCTP, and unlabeled

ddNTPs. The 5' overhang generated by digestion with BceA I and with BsmF I and the expected nucleotides are indicated.

FIG. 9C. Sequence analysis of SNP TSC0264580 after amplification with a second primer that contained the recognition site for BsmF I. Four fill in reactions are shown; each reaction contained one fluorescently labeled nucleotide, which was ddGTP, ddATP, ddTTP, or ddCTP and unlabeled ddNTPs. Two different 5' overhangs are depicted: one represents the DNA molecules that were cut 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand and the other represents the DNA molecules that were cut 10 nucleotides away on the sense strand and 14 nucleotides away on the antisense strand. The expected nucleotides also are indicated.

FIG. 9D. Sequence analysis of SNP HC21S00027 amplified with a second primer that contained the recognition site for BsmF I. A mixture of labeled ddNTPs and unlabeled dNTPs was used to fill in the 5' overhang generated by digestion with BsmF I. Two different 5' overhangs are depicted: one represents the DNA molecules that were cut 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand and the other represents the DNA molecules that were cut 10 nucleotides away on the sense strand and 14 nucleotides away on the antisense strand. The nucleotide from the SNP, the nucleotide at the SNP site (the sample contained DNA templates from 36 individuals; both nucleotides would be expected to be represented in the sample), and the three nucleotides downstream of the SNP are indicated.

FIG. 10. Sequence analysis of multiple SNPs. SNPs HC21S00131, and HC21S00027, which are located on chromosome 21, and SNPs TSC0087315, SNP TSC0214366, SNP TSC0413944, and SNP TSC0095512, which are on chromosome 1, were amplified in separate PCR reactions with second primers that contained a recognition site for BsmF I. The primers were designed so that each amplified locus of interest was of a different size. After amplification, the reactions were pooled into a single sample, and all subsequent steps of the method performed (as described for FIGS. 1F-1I) on that sample. Each SNP and the nucleotide found at each SNP are indicated.

FIG. 11. Quantification of the percentage of fetal DNA in maternal blood. Blood was obtained from a pregnant human female with informed consent. DNA was isolated and serial dilutions were made to determine the percentage of fetal DNA present in the sample. The SRY gene, which is located on chromosome Y, was used to detect fetal DNA. The cystic fibrosis gene, which is located on chromosome 7, was used to detect both maternal and fetal DNA.

FIG. 11 A. Amplification of the SRY gene and the cystic fibrosis gene using a DNA template isolated from a blood sample that was treated with EDTA.

FIG. 11B. Amplification of the SRY gene and the cystic fibrosis gene using a DNA template that was isolated from a blood sample that was treated with formalin and EDTA.

FIG. 12. Genetic analysis of an individual previously genotyped with Trisomy 21 (Down's Syndrome). Blood was collected, with informed consent, from an individual who had previously been genotyped with trisomy 21. DNA was isolated and two SNPs on chromosome 21 and two SNPs on chromosome 13 were genotyped. As shown in the photograph of the gel, the SNPs at chromosome 21 show disproportionate ratios of the two nucleotides. Visual inspection of the gel demonstrates that one nucleotide of the two nucleotides at the SNP sites analyzed for chromosome 21 is of greater intensity, suggesting it is not present in a 50:50 ratio. However, visual inspection of the gel suggests that the nucleotides at the heterozygous SNP sites analyzed on chromosome 13 are present in the expected 50:50 ratio.

FIG. 13. Sequence determination of both alleles of SNPs TSC0837969, TSC0034767, TSC1130902, TSC0597888, TSC0195492, TSC0607185 using one fluorescently labeled nucleotide. Labeled ddGTP was used in the presence of unlabeled dATP, dCTP, dTTP to fill-in the overhang generated by digestion with BsmF I. The nucleotide preceding the variable site on the strand that was filled-in was not guanine, and the nucleotide after the variable site on the strand that was filled in was not guanine. The nucleotide two bases after the variable site on the strand that was filled-in was guanine. Alleles that contain guanine at variable site are filled in with labeled ddGTP.

Alleles that do not contain guanine are filled in with unlabeled dATP, dCTP, or dTTP, and the polymerase continues to incorporate nucleotides until labeled ddGTP is filled in at position 3 complementary to the overhang.

FIG. 14. Identification of SNPs with alleles that are variable within the population. The sequences of both alleles of seven SNPs located on chromosome 13 were determined using a template DNA comprised of DNA obtained from two hundred and forty five individuals. Labeled ddGTP was used in the presence of unlabeled dATP, dCTP, dTTP to fill-in the overhang generated by digestion with BsmF I. The nucleotide preceding the variable site on the strand that was filled-in was not guanine, and the nucleotide after the variable site on the strand that was filled in was not guanine. The nucleotide two bases after the variable site on the strand that was filled-in was guanine. Alleles that contain guanine at variable site are filled in with labeled ddGTP. Alleles that do not contain guanine are filled in with unlabeled dATP, dCTP, or dTTP, and the polymerase continues to incorporate nucleotides until labeled ddGTP is filled in at position 3 complementary to the overhang.

FIG. 15. Determination of the ratio for one allele to the other allele at heterozygous SNPs. The observed nucleotides for SNP TSC0607185 are cytosine (referred to as allele 1) and thymidine (referred to as allele 2) on the sense strand. The ratio of allele 2 to allele 1 was calculated using template DNA isolated from five individuals. The ratio of allele 2 to allele 1 (allele 2 / allele 1) was consistently 1:1.

The observed nucleotides for SNP TSC1130902 are guanine (referred to as allele 1) and adenine (referred to as allele 2) on the sense strand. The ratio of allele 2 to allele 1 was calculated using template DNA isolated from five individuals. The ratio of allele 2 to allele 1 (allele 2 / allele1) was consistently 75:25.

FIG. 16. The percentage of allele 2 to allele 1 at SNP TSC0108992 remains linear when calculated on template DNA containing an extra copy of chromosome 21. SNP TSC0108992 was amplified using template DNA from four individuals, and two separate fill-in reactions (labeled as A and B) were performed for each PCR reaction (labeled 1 through 4). The calculated percentage of allele 2 to allele 1 on template DNA from normal individuals was 0.47. The deviation from the theoretically predicted

percentage of 0.50 remained linear on template DNA isolated from an individual with Down's syndrome.

FIG. 17A. Analysis of a SNP located on chromosome 21 from template DNA isolated from an individual with a normal genetic karyotype. SNP TSC0108992 was amplified using the methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.50.

FIG 17B. Analysis of a SNP located on chromosome 21 from template DNA isolated from an individual with a trisomy 21 genetic karyotype. SNP TSC0108992 was amplified using the methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.30.

FIG. 17C. Analysis of a SNP located on chromosome 21 from a mixture comprised of template DNA from an individual with Trisomy 21, and template DNA from an individual with a normal genetic karyotype in a ratio of 3:1 (Trisomy 21: Normal). SNP TSC0108992 was amplified from the mixture of template DNA using the methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.319.

FIG. 17D. Analysis of a SNP located on chromosome 21 from a mixture comprised of template DNA from an individual with Trisomy 21, and template DNA from an individual with a normal genetic karyotype in a ratio of 1:1 (Trisomy 21: Normal). SNP TSC0108992 was amplified from the mixture of template DNA using the

methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.352.

FIG. 17E. Analysis of a SNP located on chromosome 21 from a mixture comprised of template DNA from an individual with Trisomy 21, and template DNA from an individual with a normal genetic karyotype in a ratio of 1:2.3 (Trisomy 21: Normal). SNP TSC0108992 was amplified from the mixture of template DNA using the methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.382.

FIG. 17F. Analysis of a SNP located on chromosome 21 from a mixture comprised of template DNA from an individual with Trisomy 21, and template DNA from an individual with a normal genetic karyotype in a ratio of 1:4 (Trisomy 21: Normal). SNP TSC0108992 was amplified from the mixture of template DNA using the methods described herein, and after digestion with the type IIS restriction enzyme BsmF I, the 5' overhang was filled in using labeled ddTTP, and unlabeled dATP, dCTP, and dGTP. Three separate PCR reactions were performed, and each PCR reaction was split into two samples. The ratio of allele 2 to allele 1 (allele 2 / (allele 2 + allele 1)) was calculated, which resulted in mean of 0.397.

FIG. 18A. Agarose gel analysis of nine (9) SNPs amplified from template DNA. Each of the nine SNPs were amplified from genomic DNA using the methods described herein.

Lane 1 corresponds to SNP TSC0397235, lane 2 corresponds to TSC0470003, lane 3 corresponds to TSC1649726, lane 4 corresponds to TSC1261039, lane 5 corresponds to TSC0310507, lane 6 corresponds to TSC1650432, lane 7 corresponds to TSC1335008, lane 8 corresponds to TSC0128307, and lane 9 corresponds to TSC0259757.

FIG. 18B. The original template DNA was amplified using 12 base primers that annealed to various regions on chromosome 13. One hundred different primer sets were used to amplify regions throughout chromosome 13. For each of the nine SNPs, a primer that annealed approximately 130 bases from the locus of interest and 130 bases downstream of the locus of interest were used. This amplification reaction, which contained a total of 100 different primer sets, was used to amplify the regions containing the loci of interest. The resulting PCR product was used in a subsequent PCR reaction, wherein each of the nine SNPs were individually amplified using a first primer and a second primer, wherein the second primer contained the binding site for the type IIs restriction enzyme BsmF I. SNPs were loaded in the same order as FIG. 18A.

FIG. 19A. Quantification of the percentage of allele 2 to allele 1 for SNP TSC047003 on original template DNA (IA) and multiplexed template DNA (M1-M3), wherein the DNA was first amplified using 12 base primers that annealed 150 bases upstream and downstream of the loci of interest. Then, three separate PCR reactions were performed on the multiplexed template DNA, using a first and second primer.

FIG. 19B. Quantification of the percentage of allele 2 to allele 1 for SNP TSC1261039 on original template DNA (IA) and multiplexed template DNA (M1-M3), wherein the DNA was first amplified using 12 base primers that annealed 150 bases upstream and downstream of the loci of interest. Then, three separate PCR reactions were performed on the multiplexed template DNA, using a first and second primer.

FIG. 19C. Quantification of the percentage of allele 2 to allele 1 for SNP TSC310507 on original template DNA (IA) and multiplexed template DNA (M1-M3), wherein the DNA was first amplified using 12 base primers that annealed 150 bases upstream and downstream of the loci of interest. Then, three separate PCR reactions were performed on the multiplexed template DNA, using a first and second primer.

FIG. 19D. Quantification of the percentage of allele 2 to allele 1 for SNP TSC1335008 on original template DNA (IA) and multiplexed template DNA (M1-M3), wherein the DNA was first amplified using 12 base primers that annealed 150 bases

upstream and downstream of the loci of interest. Then, three separate PCR reactions were performed on the multiplexed template DNA, using a first and second primer.

FIG. 20. Detection of fetal DNA from plasma DNA isolated from a pregnant female. Four SNPs wherein the maternal DNA was homozygous were analyzed on the plasma DNA. The maternal DNA was homozygous for adenine at TSC0838335 (lane 1), while the plasma DNA displayed a heterozygous pattern (lane 2). The guanine allele represented the fetal DNA, which was clearly distinguished from the maternal signal. Both the maternal DNA and the plasma DNA were homozygous for adenine at TSC0418134 (lanes 3 and 4). The maternal DNA was homozygous for guanine at TSC0129188 (lane 5), while the plasma DNA displayed a heterozygous pattern (lane 6). The adenine allele represented the fetal DNA. Both the maternal DNA and the plasma DNA were homozygous for adenine at TSC0501389 (lanes 7 and 8).

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides a method for detecting genetic disorders, including but not limited to mutations, insertions, deletions, and chromosomal abnormalities, and is especially useful for the detection of genetic disorders of a fetus. The method is especially useful for detection of a translocation, addition, amplification, transversion, inversion, aneuploidy, polyploidy, monosomy, trisomy, trisomy 21, trisomy 13, trisomy 14, trisomy 15, trisomy 16, trisomy 18, trisomy 22, triploidy, tetraploidy, and sex chromosome abnormalities including XO, XXY, XYY, and XXX. The method also provides a non-invasive technique for determining the sequence of fetal DNA.

The invention is directed to a method for detecting chromosomal abnormalities, the method comprising: (a) determining the sequence of alleles of a locus of interest on a template DNA; and (b) quantitating a ratio for the alleles at a heterozygous locus of interest that was identified from the locus of interest of (a), wherein said ratio indicates the presence or absence of a chromosomal abnormality.

In another embodiment, the present invention provides a non-invasive method for determining the sequence of a locus of interest on fetal DNA, said method comprising: (a) obtaining a sample from a pregnant female; (b) adding a cell lysis inhibitor to the sample of (a); (c) obtaining template DNA from the sample of (b), wherein said template

DNA comprises fetal DNA and maternal DNA; and (d) determining the sequence of a locus of interest on template DNA.

DNA Template

By a "locus of interest" is intended a selected region of nucleic acid that is within a larger region of nucleic acid. A locus of interest can include but is not limited to 1-100, 1-50, 1-20, or 1-10 nucleotides, preferably 1-6, 1-5, 1-4, 1-3, 1-2, or 1 nucleotide(s).

As used herein, an "allele" is one of several alternate forms of a gene or non-coding regions of DNA that occupy the same position on a chromosome. The term allele can be used to describe DNA from any organism including but not limited to bacteria, viruses, fungi, protozoa, molds, yeasts, plants, humans, non-humans, animals, and archeabacteria.

For example, bacteria typically have one large strand of DNA. The term allele with respect to bacterial DNA refers to the form of a gene found in one cell as compared to the form of the same gene in a different bacterial cell of the same species.

Alleles can have the identical sequence or can vary by a single nucleotide or more than one nucleotide. With regard to organisms that have two copies of each chromosome, if both chromosomes have the same allele, the condition is referred to as homozygous. If the alleles at the two chromosomes are different, the condition is referred to as heterozygous. For example, if the locus of interest is SNP X on chromosome 1, and the maternal chromosome contains an adenine at SNP X (A allele) and the paternal chromosome contains a guanine at SNP X (G allele), the individual is heterozygous at SNP X.

As used herein, sequence means the identity of one nucleotide or more than one contiguous nucleotides in a polynucleotide. In the case of a single nucleotide, e.g., a SNP, "sequence" and "identity" are used interchangeably herein.

The term "chromosomal abnormality" refers to a deviation between the structure of the subject chromosome and a normal homologous chromosome. The term "normal" refers to the predominate karyotype or banding pattern found in healthy individuals of a particular species. A chromosomal abnormality can be numerical or structural, and includes but is not limited to aneuploidy, polyploidy, inversion, a trisomy, a monosomy, duplication, deletion, deletion of a part of a chromosome, addition, addition of a part of chromosome, insertion, a fragment of a chromosome, a region of a chromosome, chromosomal rearrangement, and translocation. A chromosomal abnormality can be

correlated with presence of a pathological condition or with a predisposition to develop a pathological condition. As defined herein, a single nucleotide polymorphism ("SNP") is not a chromosomal abnormality.

As used herein with respect to individuals, "mutant alleles" refers to variant alleles that are associated with a disease state.

The term "template" refers to any nucleic acid molecule that can be used for amplification in the invention. RNA or DNA that is not naturally double stranded can be made into double stranded DNA so as to be used as template DNA. Any double stranded DNA or preparation containing multiple, different double stranded DNA molecules can be used as template DNA to amplify a locus or loci of interest contained in the template DNA.

The template DNA can be obtained from any source including but not limited to humans, non-humans, mammals, reptiles, cattle, cats, dogs, goats, swine, pigs, monkeys, apes, gorillas, bulls, cows, bears, horses, sheep, poultry, mice, rats, fish, dolphins, whales, and sharks.

The template DNA can be from any appropriate sample including but not limited to, nucleic acid-containing samples of tissue, bodily fluid (for example, blood, serum, plasma, saliva, urine, tears, peritoneal fluid, ascitic fluid, vaginal secretion, lymph fluid, cerebrospinal fluid or mucosa secretion), umbilical cord blood, chorionic villi, amniotic fluid, an embryo, a two-celled embryo, a four-celled embryo, an eight-celled embryo, a 16-celled embryo, a 32- celled embryo, a 64-celled embryo, a 128-celled embryo, a 256-celled embryo, a 512-celled embryo, a 1024-celled embryo, embryonic tissues, lymph fluid, cerebrospinal fluid, mucosa secretion, or other body exudate, fecal matter, an individual cell or extract of the such sources that contain the nucleic acid of the same, and subcellular structures such as mitochondria, using protocols well established within the art.

In one embodiment, the template DNA can be obtained from a sample of a pregnant female.

In another embodiment, the template DNA can be obtained from an embryo. In a preferred embodiment, the template DNA can be obtained from a single-cell of an embryo.

In one embodiment, the template DNA is fetal DNA. Fetal DNA can be obtained from sources including but not limited to maternal blood, maternal serum, maternal plasma, fetal cells, umbilical cord blood, chorionic villi, amniotic fluid, cells or tissues.

In another embodiment, a cell lysis inhibitor is added to the sample including but not limited to formaldehyde, formaldehyde derivatives, formalin, glutaraldehyde, glutaraldehyde derivatives, primary amine reactive crosslinkers, sulfhydryl reactive crosslinkers, sulfhydryl addition or disulfide reduction, carbohydrate reactive crosslinkers, carboxyl reactive crosslinkers, photoreactive crosslinkers, cleavable crosslinkers, AEDP, APG, BASED, $BM(PEO)_3$, $BM(PEO)_4$, BMB, BMDB, BMH, BMOE, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP, DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, sulfo-BSOCOES, Sulfo-DST, or Sulfo-EGS. In another embodiment, two, three, four, five or more than five cell lysis inhibitors can be added to the sample.

In another embodiment, the template DNA contains both maternal DNA and fetal DNA. In a preferred embodiment, template DNA is obtained from blood of a pregnant female. Blood is collected using any standard technique for blood-drawing including but not limited to venipuncture. For example, blood can be drawn from a vein from the inside of the elbow or the back of the hand. Blood samples can be collected from a pregnant female at any time during fetal gestation. For example, blood samples can be collected from human females at 1-4, 4-8, 8-12, 12-16, 16-20, 20-24, 24-28, 28-32, 32-36, 36-40, or 40-44 weeks of fetal gestation, and preferably between 8-28 weeks of fetal gestation.

The blood sample is centrifuged to separate the plasma from the maternal cells. The plasma and maternal cell fractions are transferred to separate tubes and re-centrifuged. The plasma fraction contains cell-free fetal DNA and maternal DNA. Any standard DNA isolation technique can be used to isolate the fetal DNA and the maternal DNA including but not limited to QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183).

In a preferred embodiment, blood can be collected into an apparatus containing a magnesium chelator including but not limited to EDTA, and is stored at 4°C. Optionally, a calcium chelator, including but not limited to EGTA, can be added.

In another embodiment, a cell lysis inhibitor is added to the maternal blood including but not limited to formaldehyde, formaldehyde derivatives, formalin,

glutaraldehyde, glutaraldehyde derivatives, primary amine reactive crosslinkers, sulfhydryl reactive crosslinkers, sulfydryl addition or disulfide reduction, carbohydrate reactive crosslinkers, carboxyl reactive crosslinkers, photoreactive crosslinkers, cleavable crosslinkers, AEDP, APG, BASED, $BM(PEO)_3$, $BM(PEO)_4$, BMB, BMDB, BMH, BMOE, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP, DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, sulfo-BSOCOES, Sulfo-DST, or Sulfo-EGS.

In another embodiment, the template DNA is obtained from the plasma or serum of the blood of the pregnant female. The percentage of fetal DNA in maternal plasma is between 0.39-11.9% (*Pertl, and Bianchi, Obstetrics and Gynecology 98*: 483-490 (2001)). The majority of the DNA in the plasma sample is maternal, which makes using the DNA for genotyping the fetus difficult. However, methods that increase the percentage of fetal DNA in the maternal plasma allow the sequence of the fetal DNA to be determined, and allow for the detection of genetic disorders including mutations, insertions, deletions, and chromosomal abnormalities. The addition of cell lysis inhibitors to the maternal blood sample can increase the relative percentage of fetal DNA. While lysis of both maternal and fetal cells is inhibited, the vast majority of cells are maternal, and thus by reducing the lysis of maternal cells, there is a relative increase in the percentage of free fetal DNA. See Example 4.

In another embodiment, any blood drawing technique, method, protocol, or equipment that reduce the amount of cell lysis can be used, including but not limited to a large boar needle, a shorter length needle, a needle coating that increases laminar flow, e.g., teflon, a modification of the bevel of the needle to increase laminar flow, or techniques that reduce the rate of blood flow. The fetal cells likely are destroyed in the maternal blood by the mother's immune system. However, it is likely that a large portion of the maternal cell lysis occurs as a result of the blood draw. Thus, methods that prevent or reduce cell lysis will reduce the amount of maternal DNA in the sample, and increase the relative percentage of free fetal DNA.

In another embodiment, an agent that preserves the structural integrity of cells can be used to reduce the amount of cell lysis.

In another embodiment, agents that prevent the destruction of DNA, including but not limited to a DNase inhibitor, zinc chloride, ethylenediaminetetraacetic acid, guanidine-HCl, guanidine isothiocyanate, N-lauroylsarcosine, and Na-dodecylsulphate, can be added to the blood sample.

In another embodiment, fetal DNA is obtained from a fetal cell, wherein said fetal cell can be isolated from sources including but not limited to maternal blood, umbilical cord blood, chorionic villi, amniotic fluid, embryonic tissues and mucous obtained from the cervix or vagina of the mother.

In a preferred embodiment, fetal cells are isolated from maternal peripheral blood. An antibody specific for fetal cells can be used to purify the fetal cells from the maternal serum (Mueller et al., Lancet 336: 197-200 (1990); Ganshirt-Ahlert *et al., Am. J. Obstet. Gynecol. 166*: 1350-1355 (1992)). Flow cytometry techniques can also be used to enrich fetal cells (Herzenberg et al., *PNAS 76*: 1453-1455 (1979); Bianchi et al., *PNAS 87*: 3279-3283 (1990); Bruch *et al., Prenatal Diagnosis* 11: 787-798 (1991)). U.S. Pat. No. 5,432,054 also describes a technique for separation of fetal nucleated red blood cells, using a tube having a wide top and a narrow, capillary bottom made of polyethylene. Centrifugation using a variable speed program results in a stacking of red blood cells in the capillary based on the density of the molecules. The density fraction containing low density red blood cells, including fetal red blood cells, is recovered and then differentially hemolyzed to preferentially destroy maternal red blood cells. A density gradient in a hypertonic medium is used to separate red blood cells, now enriched in the fetal red blood cells from lymphocytes and ruptured maternal cells. The use of a hypertonic solution shrinks the red blood cells, which increases their density, and facilitate purification from the more dense lymphocytes. After the fetal cells have been isolated, fetal DNA can be purified using standard techniques in the art.

The nucleic acid that is to be analyzed can be any nucleic acid, e.g., genomic, plasmid, cosmid, yeast artificial chromosomes, artificial or man-made DNA, including unique DNA sequences, and also DNA that has been reverse transcribed from an RNA sample, such as cDNA. The sequence of RNA can be determined according to the invention if it is capable of being made into a double stranded DNA form to be used as template DNA.

The terms "primer" and "oligonucleotide primer" are interchangeable when used to discuss an oligonucleotide that anneals to a template and can be used to prime the synthesis of a copy of that template.

"Amplified" DNA is DNA that has been "copied" once or multiple times, e.g. by polymerase chain reaction. When a large amount of DNA is available to assay, such that a sufficient number of copies of the locus of interest are already present in the sample to

be assayed, it may not be necessary to "amplify" the DNA of the locus of interest into an even larger number of replicate copies. Rather, simply "copying" the template DNA once using a set of appropriate primers, which may contain hairpin structures that allow the restriction enzyme recognition sites to be double stranded, can suffice.

"Copy" as in "copied DNA" refers to DNA that has been copied once, or DNA that has been amplified into more than one copy.

In one embodiment, the nucleic acid is amplified directly in the original sample containing the source of nucleic acid. It is not essential that the nucleic acid be extracted, purified or isolated; it only needs to be provided in a form that is capable of being amplified. Hybridization of the nucleic acid template with primer, prior to amplification, is not required. For example, amplification can be performed in a cell or sample lysate using standard protocols well known in the art. DNA that is on a solid support, in a fixed biological preparation, or otherwise in a composition that contains non-DNA substances and that can be amplified without first being extracted from the solid support or fixed preparation or non-DNA substances in the composition can be used directly, without further purification, as long as the DNA can anneal with appropriate primers, and be copied, especially amplified, and the copied or amplified products can be recovered and utilized as described herein.

In a preferred embodiment, the nucleic acid is extracted, purified or isolated from non-nucleic acid materials that are in the original sample using methods known in the art prior to amplification.

In another embodiment, the nucleic acid is extracted, purified or isolated from the original sample containing the source of nucleic acid and prior to amplification, the nucleic acid is fragmented using any number of methods well known in the art including but not limited to enzymatic digestion, manual shearing, and sonication. For example, the DNA can be digested with one or more restriction enzymes that have a recognition site, and especially an eight base or six base pair recognition site, which is not present in the loci of interest. Typically, DNA can be fragmented to any desired length, including 50, 100, 250, 500, 1,000, 5,000, 10,000, 50,000 and 100,000 base pairs long. In another embodiment, the DNA is fragmented to an average length of about 1000 to 2000 base pairs. However, it is not necessary that the DNA be fragmented.

Fragments of DNA that contain the loci of interest can be purified from the fragmented DNA before amplification. Such fragments can be purified by using primers

that will be used in the amplification (see "Primer Design" section below) as hooks to retrieve the loci of interest, based on the ability of such primers to anneal to the loci of interest. In a preferred embodiment, tag-modified primers are used, such as e.g. biotinylated primers.

By purifying the DNA fragments containing the loci of interest, the specificity of the amplification reaction can be improved. This will minimize amplification of nonspecific regions of the template DNA. Purification of the DNA fragments can also allow multiplex PCR (Polymerase Chain Reaction) or amplification of multiple loci of interest with improved specificity.

The loci of interest that are to be sequenced can be selected based upon sequence alone. In humans, over 1.42 million single nucleotide polymorphisms (SNPs) have been described (*Nature 409*:928-933 (2001); The SNP Consortium LTD). On the average, there is one SNP every 1.9 kb of human genome. However, the distance between loci of interest need not be considered when selecting the loci of interest to be sequenced according to the invention. If more than one locus of interest on genomic DNA is being analyzed, the selected loci of interest can be on the same chromosome or on different chromosomes.

In a preferred embodiment, the selected loci of interest can be clustered to a particular region on a chromosome. Multiple loci of interest can be located within a region of DNA such that even with any breakage or fragmentation of the DNA, the multiple loci of interest remain linked. For example, if the DNA is obtained and by natural forces is broken into fragments of 5 Kb, multiple loci of interest can be selected within the 5 Kb regions. This allows each fragment, as measured by the loci of interest within that fragment, to serve as an experimental unit, and will reduce any possible experimental noise of comparing loci of interest on multiple chromosomes.

The loci of interest on a chromosome can be any distance from each other including but not limited to 10-50, 50-100, 100-150, 150-200, 200-250, 250-500, 500-750, 750-1000, 1000-1500, 1500-2000, 2000-2500, 2500-3000, 3000-3500, 3500-4000, 4000-4500, 4500-5000, 5000-10,000 and greater than 10,000 base pairs.

In a preferred embodiment, the length of sequence that is amplified is preferably different for each locus of interest so that the loci of interest can be separated by size.

In fact, it is an advantage of the invention that primers that copy an entire gene sequence need not be utilized. Rather, the copied locus of interest is preferably only a

small part of the total gene or a small part of a non-coding region of DNA. There is no advantage to sequencing the entire gene as this can increase cost and delay results. Sequencing only the desired bases or loci of interest maximizes the overall efficiency of the method because it allows for the sequence of the maximum number of loci of interest to be determined in the fastest amount of time and with minimal cost.

Because a large number of sequences can be analyzed together, the method of the invention is especially amenable to the large-scale screening of a number of loci of interest.

Any number of loci of interest can be analyzed and processed, especially at the same time, using the method of the invention. The sample(s) can be analyzed to determine the sequence at one locus of interest or at multiple loci of interest at the same time. The loci of interest can be present on a single chromosome or on multiple chromosomes.

Alternatively, 2, 3, 4, 5, 6, 7, 8, 9, 10-20, 20-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-100, 100-250, 250-500, 500-1,000, 1,000-2,000, 2,000-3, 000, 3,000-5,000, 5,000-10,000, 10,000-50,000 or more than 50,000 loci of interest can be analyzed at the same time when a global genetic screening is desired. Such a global genetic screening might be desired when using the method of the invention to provide a genetic fingerprint to identify an individual or for SNP genotyping.

The locus of interest that is to be copied can be within a coding sequence or outside of a coding sequence. Preferably, one or more loci of interest that are to be copied are within a gene. In a preferred embodiment, the template DNA that is copied is a locus or loci of interest that is within a genomic coding sequence, either intron or exon. In a highly preferred embodiment, exon DNA sequences are copied. The loci of interest can be sites where mutations are known to cause disease or predispose to a disease state. The loci of interest can be sites of single nucleotide polymorphisms. Alternatively, the loci of interest that are to be copied can be outside of the coding sequence, for example, in a transcriptional regulatory region, and especially a promoter, enhancer, or repressor sequence.

Method for Determining the Sequence of a Locus of Interest

Any method that provides information on the sequence of a nucleic acid can be used including but not limited to allele specific PCR, PCR, mass spectrometry, MALDI-TOF mass spectrometry hybridization, primer extension, fluorescence detection,

fluorescence resonance energy transfer (FRET), fluorescence polarization, DNA sequencing, Sanger dideoxy sequencing, DNA sequencing gels, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, microarray, southern blot, slot blot, dot blot, and single primer linear nucleic acid amplification, as described in U.S. Patent No. 6,251,639.

The preferred method of determining the sequence has previously been described in U.S. Application No. 10/093,618, filed on March 11, 2002, hereby incorporated by reference in its entirety.

I. Primer Design

Published sequences, including consensus sequences, can be used to design or select primers for use in amplification of template DNA. The selection of sequences to be used for the construction of primers that flank a locus of interest can be made by examination of the sequence of the loci of interest, or immediately thereto. The recently published sequence of the human genome provides a source of useful consensus sequence information from which to design primers to flank a desired human gene locus of interest.

By "flanking" a locus of interest is meant that the sequences of the primers are such that at least a portion of the 3' region of one primer is complementary to the antisense strand of the template DNA and from the locus of interest site (forward primer), and at least a portion of the 3' region of the other primer is complementary to the sense strand of the template DNA and downstream of the locus of interest (reverse primer). A "primer pair" is intended a pair of forward and reverse primers. Both primers of a primer pair anneal in a manner that allows extension of the primers, such that the extension results in amplifying the template DNA in the region of the locus of interest.

Primers can be prepared by a variety of methods including but not limited to cloning of appropriate sequences and direct chemical synthesis using methods well known in the art (Narang *et al., Methods Enzymol. 68*:90 (1979); Brown *et al., Methods Enzymol. 68*:109 (1979)). Primers can also be obtained from commercial sources such as Operon Technologies, Amersham Pharmacia Biotech, Sigma, and Life Technologies. The primers can have an identical melting temperature. The lengths of the primers can be extended or shortened at the 5' end or the 3' end to produce primers with desired melting temperatures. In a preferred embodiment, one of the primers of the prime pair is longer than the other primer. In a preferred embodiment, the 3' annealing lengths of the primers, within a primer pair, differ. Also, the annealing position of each primer pair can be

designed such that the sequence and length of the primer pairs yield the desired melting temperature. The simplest equation for determining the melting temperature of primers smaller than 25 base pairs is the Wallace Rule (Td = 2(A+T) + 4(G+C)). Computer programs can also be used to design primers, including but not limited to Array Designer Software (Arrayit Inc.), Oligonucleotide Probe Sequence Design Software for Genetic Analysis (Olympus Optical Co.), NetPrimer, and DNAsis from Hitachi Software Engineering. The TM (melting or annealing temperature) of each primer is calculated using software programs such as Net Primer (free web based program at http://premierbiosoft.com/netprimer/netprlaunch/netprlaunch.html; internet address as of April 17, 2002).

In another embodiment, the annealing temperature of the primers can be recalculated and increased after any cycle of amplification, including but not limited to cycle 1, 2, 3, 4, 5, cycles 6-10, cycles 10-15, cycles 15-20, cycles 20-25, cycles 25-30, cycles 30-35, or cycles 35-40. After the initial cycles of amplification, the 5' half of the primers is incorporated into the products from each loci of interest, thus the TM can be recalculated based on both the sequences of the 5' half and the 3' half of each primer.

For example, in FIG. 1B, the first cycle of amplification is performed at about the melting temperature of the 3' region, which anneals to the template DNA, of the second primer (region "c"), which is 13 bases. After the first cycle, the annealing temperature can be raised to TM2, which is about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer, which is depicted as region "b." The second primer cannot bind to the original template DNA because it only anneals to 13 bases in the original DNA template, and TM2 is about the melting temperature of approximately 20 bases, which is the 3' annealing region of the first primer (FIG. 1C). However, the first primer can bind to the DNA that was copied in the first cycle of the reaction. In the third cycle, the annealing temperature is raised to TM3, which is about the melting temperature of the entire sequence of the second primer, which is depicted as regions "c" and "d." The DNA template produced from the second cycle of PCR contains both regions c' and d', and therefore, the second primer can anneal and extend at TM3 (FIG. 1D). The remaining cycles are performed at TM3. The entire sequence of the first primer (a + b') can anneal to the template from the third cycle of PCR, and extend (FIG. 1E). Increasing the annealing temperature will decrease non-specific binding and increase the specificity of the reaction, which is especially

useful if amplifying a locus of interest from human genomic DNA, which is about $3x10^9$ base pairs long.

As used herein, the term "about" with regard to annealing temperatures is used to encompass temperatures within 10 degrees celcius of the stated temperatures.

In one embodiment, one primer pair is used for each locus of interest. However, multiple primer pairs can be used for each locus of interest.

In one embodiment, primers are designed such that one or both primers of the primer pair contain sequence in the 5' region for one or more restriction endonucleases (restriction enzyme).

As used herein, with regard to the position at which restriction enzymes digest DNA, the "sense" strand is the strand reading 5' to 3' in the direction in which the restriction enzyme cuts. For example, BsmF I recognizes the following sequences:

5' GGGAC$(N)_{10}$ 3'	5' $(N)_{14}$GT000 3'
3' CCCTG$(N)_{14}$ 5'	3'$(N)_{10}$CAGGG 5'

The sense strand is the strand containing the "GGGAC" sequence as it reads 5' to 3' in the direction that the restriction enzyme cuts.

As used herein, with regard to the position at which restriction enzymes digest DNA, the "antisense" strand is the strand reading 3' to 5' in the direction in which the restriction enzyme cuts.

In another embodiment, one of the primers in a primer pair is designed such that it contains a restriction enzyme recognition site for a restriction enzyme that cuts "n" nucleotides away from the recognition site, and produces a recessed 3' end and a 5' overhang that contains the locus of interest (herein referred to as a "second primer"). "N" is a distance from the recognition site to the site of the cut by the restriction enzyme. In other words, the second primer of a primer pair contains a recognition site for a restriction enzyme that does not cut DNA at the recognition site but cuts "n" nucleotides away from the recognition site. For example, if the recognition sequence is for the restriction enzyme BceA I, the enzyme will cut ten (10) nucleotides from the recognition site on the sense strand, and twelve (12) nucleotides away from the recognition site on the antisense strand.

The 3' region and preferably, the 3' half, of the primers is designed to anneal to a sequence that flanks the loci of interest (FIG. 1A). The second primer can anneal any distance from the locus of interest provided that digestion with the restriction enzyme that

recognizes the restriction enzyme recognition site on this primer generates a 5' overhang that contains the locus of interest. The 5' overhangs can be of any size, including but not limited to 1, 2, 3, 4, 5, 6, 7, 8, and more than 8 bases.

In a preferred embodiment, the 3' end of the primer that anneals closer to the locus of interest (second primer) can anneal 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or more than 14 bases from the locus of interest or at the locus of interest.

In a preferred embodiment, the second primer is designed to anneal closer to the locus of interest than the other primer of a primer pair (the other primer is herein referred to as a "first primer"). The second primer can be a forward or reverse primer and the first primer can be a reverse or forward primer, respectively. Whether the first or second primer should be the forward or reverse primer can be determined by which design will provide better sequencing results.

For example, the primer that anneals closer to the locus of interest can contain a recognition site for the restriction enzyme BsmF I, which cuts ten (10) nucleotides from the recognition site on the sense strand, and fourteen (14) nucleotides from the recognition site on the antisense strand. In this case, the primer can be designed so that the restriction enzyme recognition site is 13 bases, 12 bases, 10 bases or 11 bases from the locus of interest. If the recognition site is 13 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (RXXX), wherein the locus of interest (R) is the first nucleotide in the overhang (reading 3' to 5'), and X is any nucleotide. If the recognition site is 12 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (XRXX), wherein the locus of interest (R) is the second nucleotide in the overhang (reading 3' to 5'). If the recognition site is 11 bases from the locus of interest, digestion with BsmF I will generate a 5' overhang (XXRX), wherein the locus of interest (R) is the third nucleotide in the overhang (reading 3' to 5'). The distance between the restriction enzyme recognition site and the locus of interest should be designed so that digestion with the restriction enzyme generates a 5' overhang, which contains the locus of interest. The effective distance between the recognition site and the locus of interest will vary depending on the choice of restriction enzyme.

In another embodiment, the primer that anneals closer to the locus of interest site, relative to the other primer, can be designed so that the restriction enzyme that generates the 5' overhang, which contains the locus of interest, will see the same sequence at the cut site, independent of the nucleotide at the locus of interest site. For example, if the

primer that anneals closer to the locus of interest is designed so that the recognition site for the restriction enzyme BsmF I (5' GGGAC 3') is thirteen bases from the locus of interest, the restriction enzyme will cut the antisense strand one base from the locus of interest. The nucleotide at the locus of interest is adjacent to the cut site, and may vary from DNA molecule to DNA molecule. If it is desired that the nucleotides adjacent to the cut site be identical, the primer can be designed so that the restriction enzyme recognition site for BsmF I is twelve bases away from the locus of interest site. Digestion with BsmF I will generate a 5' overhang, wherein the locus of interest site is in the second position of the overhang (reading 3' to 5') and is no longer adjacent to the cut site. Designing the primer so that the restriction enzyme recognition site is twelve (12) bases from the locus of interest site allows the nucleotides adjacent to the cut site to be the same, independent of the nucleotide at the locus of interest. Also, primers that have been designed so that the restriction enzyme recognition site, BsmF I, is eleven (11) or ten (10) bases from the locus of interest site will allow the nucleotides adjacent to the cut site to be the same, independent of the nucleotide at the locus of interest. Similar strategies of primer design can be employed with other restriction enzymes so that the nucleotides adjacent to the cut site will be the same, independent of the nucleotide at the loci of interest.

The 3' end of the first primer (either the forward or the reverse) can be designed to anneal at a chosen distance from the locus of interest. Preferably, for example, this distance is between 10-25, 25-50, 50-75, 75-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-550, 550-600, 600-650, 650-700, 700-750, 750-800, 800-850, 850-900, 900-950, 950-1000 and greater than 1000 bases away from the locus of interest. The annealing. sites of the first primers are chosen such that each successive upstream primer is further and further away from its respective downstream primer.

For example, if at locus of interest 1 the 3' ends of the first and second primers are Z bases apart, then at locus of interest 2, the 3' ends of the upstream and downstream primers are Z + K bases apart, where K = 1, 2, 3, 4, 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, or greater than 1000 bases (FIG 2). The purpose of making the first primers further and further apart from their respective second primers is so that the PCR products of all the loci of interest differ in size and can be separated,

e.g., on a sequencing gel. This allows for multiplexing by pooling the PCR products in later steps.

In one embodiment, the 5' region of the first or second primer can have a recognition site for any type of restriction enzyme. In a preferred embodiment, the 5' region of the first and/or second primer has at least one restriction enzyme recognition site that is different from the restriction enzyme recognition site that is used to generate the 5' overhang, which contains the locus of interest.

In one embodiment, the 5' region of the first primer can have a recognition site for any type of restriction enzyme. In a preferred embodiment, the first primer has at least one restriction enzyme recognition site that is different from the restriction enzyme recognition site in the second primer. In another preferred embodiment, the first primer anneals further away from the locus of interest than the second primer.

In a preferred embodiment, the second primer contains a restriction enzyme recognition sequence for a Type IIS restriction enzyme including but not limited to BceA I and BsmF I, which produce a two base 5' overhang and a four base 5' overhang, respectively. Restriction enzymes that are Type IIS are preferred because they recognize asymmetric base sequences (not palindromic like the orthodox Type II enzymes). Type IIS restriction enzymes cleave DNA at a specified position that is outside of the recognition site, typically up to 20 base pairs outside of the recognition site. These properties make Type IIS restriction enzymes, and the recognition sites thereof, especially useful in the method of the invention. Preferably, the Type IIS restriction enzymes used in this method leave a 5' overhang and a recessed 3'.

A wide variety of Type IIS restriction enzymes are known and such enzymes have been isolated from bacteria, phage, archeabacteria and viruses of eukaryotic algae and are commercially available (Promega, Madison WI; New England Biolabs, Beverly, MA; Szybalski W. et al., Gene 100:13-26, 1991). Examples of Type IIS restriction enzymes that would be useful in the method of the invention include, but are not limited to enzymes such as those listed in Table I.

Enzyme-Source	Recognition/Cleavage Site	Supplier
Alw I - *Acinetobacter lwoffii*	GGATC(4/5)	NE Biolabs
Alw26 I - *Acinetobacter lwoffi*	GTCTC(1/5)	Promega
Bbs I - *Bacillus laterosporus*	GAAGAC(2/6)	NE Biolabs

Bbv I - *Bacillus brevis*	GCAGC(8/12)	NE Biolabs
BceA I - *Bacillus cereus 1315*	IACGGC(12/14)	NE Biolabs
Bmr I - *Bacillus megaterium*	CTGGG(5/4)	NE Biolabs
Bsa I - *Bacillus stearothermophilus 6-55*	GGTCTC(1/5)	NE Biolabs
Bst71 I - *Bacillus stearothermophilus 71*	GCAGC(8/12)	Promega
BsmA I - *Bacillus stearothermophilus A664*	GTCTC(1/5)	NE Biolabs
BsmB I - *Bacillus stearothermophilus B61*	CGTCTC(1/5)	NE Biolabs
BsmF I - *Bacillus stearothermophilus F*	GGGAC(10/14)	NE Biolabs
BspM I - *Bacillus species M*	ACCTGC(4/8)	NE Biolabs
Ear I - *Enterobacter aerogenes*	CTCTTC(1/4)	NE Biolabs
Fau I - *Flavobacterium aquatile*	CCCGC(4/6)	NE Biolabs
Fok I - *Flavobacterium okeonokoites*	GGATG(9/13)	NE Biolabs
Hga I - *Haemophilus gallinarum*	GACGC(5/10)	NE Biolabs
Ple I - Pseudomonas lemoignei	GAGTC(4/5)	NE Biolabs
Sap I - *Saccharopolyspora species*	GCTCTTC(1/4)	NE Biolabs
SfaN I - *Streptococcus faecalis ND547*	GCATC(5/9)	NE Biolabs
Sth132 I - *Streptococcus thermophilus ST132*	CCCG(4/8)	No commercial supplier (Gene 195:201-206 (1997))

In one embodiment, a primer pair has sequence at the 5' region of each of the primers that provides a restriction enzyme recognition site that is unique for one restriction enzyme.

In another embodiment, a primer pair has sequence at the 5' region of each of the primers that provide a restriction site that is recognized by more than one restriction enzyme, and especially for more than one Type IIS restriction enzyme. For example, certain consensus sequences can be recognized by more than one enzyme. For example, BsgI, Eco571 and BpmI all recognize the consensus (G/C)TGnAG and cleave 16 by away on the antisense strand and 14 by away on the sense strand. A primer that provides such a consensus sequence would result in a product that has a site that can be recognized by any of the restriction enzymes BsgI, Eco571 and BpmI.

Other restriction enzymes that cut DNA at a distance from the recognition site, and produce a recessed 3' end and a 5' overhang include Type III restriction enzymes.

For example, the restriction enzyme EcoP15I recognizes the sequence 5' CAGCAG 3' and cleaves 25 bases downstream on the sense strand and 27 bases on the antisense strand. It will be further appreciated by a person of ordinary skill in the art that new restriction enzymes are continually being discovered and can readily be adopted for use in the subject invention.

In another embodiment, the second primer can contain a portion of the recognition sequence for a restriction enzyme, wherein the full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest. For example, the recognition site for BsmF I is 5' $GGGAC N_{10}^{\downarrow}$ 3'. The 3' region, which anneals to the template DNA, of the second primer can end with the nucleotides "GGG," which do not have to be complementary with the template DNA. If the 3' annealing region is about 10-20 bases, even if the last three bases do not anneal, the primer will extend and, generate a BsmF I site.

Second primer: 5' GGAAATTCCATGATGCGTGGG→

Template DNA 3' CCTTTAAGGTACTACGCA$N_1N_2N_3$TG 5'

5' GGAAATTCCATGATGCCT$N_1N_2N_3$AC 3'

The second primer can be designed to anneal to the template DNA, wherein the next two bases of the template DNA are thymidine and guanine, such that an adenosine and cytosine are incorporated into the primer forming a recognition site for BsmF I, 5' $GGGAC N_{10}^{\downarrow}$ 3'. The second primer can be designed to anneal in such a manner that digestion with BsmF I generates a 5' overhang containing the locus of interest.

In another embodiment, the second primer can contain an entire or full recognition site for a restriction enzyme or a portion of a recognition site, which generates a full recognition site upon primer-dependent replication of the template DNA such that digestion with a restriction enzyme that cuts at the recognition site and generates a 5' overhang that contains the locus of interest. For example, the restriction enzyme BsaJ I binds the following recognition site: 5' $C^{\downarrow}CN_1N_2GG$ 3'. The second primer can be designed such that the 3' region, which anneals to the template DNA of the primer ends with "CC", the SNP of interest is represented by "N_1", and the template sequence downstream of the SNP is "N_2GG."

Second primer: 5' GGAAATTCCATGATGCGTACC→

Template DNA 3' CCTTTAAGGTACTACGCATGGN_1N_2CC 5'

5' GGAAATTCCATGATGCCTACCN_1N_2GG 3'

After digestion with BsaJ I, a 5' overhang of the following sequence would be generated:

5' C 3'

3' GGN_1N_2CC 5'

If the nucleotide guanine is not reported at the locus of interest, the 3' recessed end can be filled in with unlabeled cytosine, which is complementary to the first nucleotide in the overhang. After removing the excess cytosine, labeled ddNTPs can be used to fill in the next nucleotide, N_1, which represents the locus of interest. Other restriction enzymes can be used including but not limited to BssK I (5' $^{\downarrow}$CCNGG 3'), Dde I(5' C$^{\downarrow}$TNAG 3'), EcoN I (5' CCTNN$^{\downarrow}$NNNAGG 3'), Fnu4H I (5' GC$^{\downarrow}$NGC 3'), Hinf I (5' G$^{\downarrow}$ANTC 3') PflF I (5' GACN$^{\downarrow}$NNGTC 3'), Sau96 I(5' G$^{\downarrow}$GNCC 3'), ScrF I (5' CC$^{\downarrow}$NGG 3'), and Tth1 11 I (5' GACN$^{\downarrow}$NNGTC 3').

It is not necessary that the 3' region, which anneals to the template DNA, of the second primer be 100% complementary to the template DNA. For example, the last 1, 2, or 3 nucleotides of the 3' end of the second primer can be mismatches with the template DNA. The region of the primer that anneals to the template DNA will target the primer, and allow the primer to extend. Even if the last two nucleotides are not complementary to the template DNA, the primer will extend and generate a restriction enzyme recognition site. For example, the last two nucleotides in the second primer are "CC." The second primer anneals to the template DNA, and allows extension even if "CC" is not complementary to the nucleotides Na, and Nb, on the template DNA.

Second primer: 5' GGAAATTCCATGATGCGTACC→

Template DNA 3' CCTTTAAGGTACTACGCATN_a·N_b·N_1·N_2·CC 5'

5' GGAAATTCCATGATGCCTA$N_aN_bN_1N_2$GG 3'

After digestion with BsaJ I, a 5' overhang of the following sequence would be generated:

5' C 3'

3' GGN_1N_2CC 5'

If the nucleotide guanine is not reported at the locus of interest, the 5' overhang can be filled in with unlabeled cytosine. The excess cytosine can be rinsed away, and

filled in with labeled ddNTPs. The first nucleotide incorporated (N_1') corresponds to the locus of interest. If guanine is reported at the locus of interest, the loci of interest can be filled in with unlabeled cytosine and a nucleotide downstream of the locus of interest can be detected. For example, assume N_2 is adenine. If the locus of interest is guanine, unlabeled cytosine can be used in the fill in reaction. After removing the cytosine, a fill in reaction with labeled thymidine can be used. The labeled thymidine will be incorporated only if the locus of interest was a guanine. Thus, the sequence of the locus of interest can be determined by detecting a nucleotide downstream of the locus of interest.

In another embodiment, the first and second primers contain a portion of a recognition sequence for a restriction enzyme, wherein the full recognition site for the restriction enzyme is generated upon amplification of the template DNA such that digestion with the restriction enzyme generates a 5' overhang containing the locus of interest. The recognition site for any restriction enzyme that contains one or more than one variable nucleotide can be generated including but not limited to the restriction enzymes BssK I (5'$^{\downarrow}$CCNGG 3'), Dde I (5'C$^{\downarrow}$TNAG 3'), Econ I (5'CCTNN$^{\downarrow}$NNNAGG 3'), Fnu4H I (5'GC$^{\downarrow}$NGC 3'), Hinf I (5'G$^{\downarrow}$ANTC 3'), PflF I (5' GACN$^{\downarrow}$NNGTC 3'), Sau96 I (5' G$^{\downarrow}$GNCC 3'), ScrF I (5' CC$^{\downarrow}$NGG 3'), and Tth1 11 I (5' GACN$^{\downarrow}$NNGTC 3').

In a preferred embodiment, the 3' regions of the first and second primers contain the partial sequence for a restriction enzyme, wherein the partial sequence contains 1, 2, 3, 4 or more than 4 mismatches with the template DNA; these mismatches create the restriction enzyme recognition site. The number of mismatches that can be tolerated at the 3' end depends on the length of the primer. For example, if the locus of interest is represented by N_1, a first primer can be designed to be complementary to the template DNA, depicted below as region "a." The 3' region of the first primer ends with "CC," which is not complementary to the template DNA. The second primer is designed to be complementary to the template DNA, which is depicted below as region "b' ". The 3' region of the second primer ends with "CC," which is not complementary to the template DNA.

First primer	5'	a	CC→		
Template DNA	3'	a'	AAN_1N_2TT	b'	5'
	5'	a	TTN_1N_2AA	b	3'

←CC b' 5' Second primer

After one round of amplification the following products would be generated:

5' a CCN_1N_2AA b 3'

and

5' b CCN_2N_1AA a' 3'.

In cycle two, the primers can anneal to the templates that were generated from the first cycle of PCR:

5' a CCN_1N_2AA b 3'
←CC b' 5'
←CC a 5'
5' b' CCN_2N_1AA a' 3'

After cycle two of PCR, the following products would be generated:

5' a CCN_1N_2GG b 3'

3' a' GGN_1N_2CC b' 5'

The restriction enzyme recognition site for BsaJ I is generated, and after digestion with BsaJ I, a 5' overhang containing the locus of interest is created. The locus of interest can be detected as described in detail below.

In another embodiment, a primer pair has sequence at the 5' region of each of the primers that provides two or more restriction sites that are recognized by two or more restriction enzymes.

In a most preferred embodiment, a primer pair has different restriction enzyme recognition sites at the 5' regions, especially 5' ends, such that a different restriction enzyme is required to cleave away any undesired sequences. For example, the first primer for locus of interest "A" can contain sequence recognized by a restriction enzyme, "X," which can be any type of restriction enzyme, and the second primer for locus of interest "A," which anneals closer to the locus of interest, can contain sequence for a restriction enzyme, "Y," which is a Type IIS restriction enzyme that cuts "n" nucleotides

away and leaves a 5'overhang and a recessed 3' end. The 5' overhang contains the locus of interest. After binding the amplified DNA to streptavidin coated wells, one can digest with enzyme "Y," rinse, then fill in with labeled nucleotides and rinse, and then digest with restriction enzyme "X," which will release the DNA fragment containing the locus of interest from the solid matrix. The locus of interest can be analyzed by detecting the labeled nucleotide that was "filled in" at the locus of interest, e.g. SNP site.

In another embodiment, the second primers for the different loci of interest that are being amplified according to the invention contain recognition sequence in the 5' regions for the same restriction enzyme and likewise all the first primers also contain the same restriction enzyme recognition site, which is a different enzyme from the enzyme that recognizes the second primers.

In another embodiment, the second primers for the multiple loci of interest that are being amplified according to the invention contain restriction enzyme recognition sequences in the 5' regions for different restriction enzymes.

In another embodiment, the first primers for the multiple loci of interest that are being amplified according to the invention contain restriction enzyme recognition sequences in the 5' regions for different restriction enzymes. Multiple restriction enzyme sequences provide an opportunity to influence the order in which pooled loci of interest are released from the solid support. For example, if 50 loci of interest are amplified, the first primers can have a tag at the extreme 5' end to aid in purification and a restriction enzyme recognition site, and the second primers can contain a recognition site for a type IIS restriction enzyme. For example, several of the first primers can have a restriction enzyme recognition site for EcoR I, other first primers can have a recognition site for Pst I, and still other first primers can have a recognition site for BamH I. After amplification, the loci of interest can be bound to a solid support with the aid of the tag on the first primers. By performing the restriction digests one restriction enzyme at a time, one can serially release the amplified loci of interest. If the first digest is performed with EcoR I, the loci of interest amplified with the first primers containing the recognition site for EcoR I will be released, and collected while the other loci of interest remain bound to the solid support. The amplified loci of interest can be selectively released from the solid support by digesting with one restriction enzyme at a time. The use of different restriction enzyme recognition sites in the first primers allows a larger number of loci of interest to be amplified in a single reaction tube.

In a preferred embodiment, any region 5' of the restriction enzyme digestion site of each primer can be modified with a functional group that provides for fragment manipulation, processing, identification, and/or purification. Examples of such functional groups, or tags, include but are not limited to biotin, derivatives of biotin, carbohydrates, haptens, dyes, radioactive molecules, antibodies, and fragments of antibodies, peptides, and immunogenic molecules.

In another embodiment, the template DNA can be replicated once, without being amplified beyond a single round of replication. This is useful when there is a large amount of the DNA available for analysis such that a large number of copies of the loci of interest are already present in the sample, and further copies are not needed. In this embodiment, the primers are preferably designed to contain a "hairpin" structure in the 5' region, such that the sequence doubles back and anneals to a sequence internal to itself in a complementary manner. When the template DNA is replicated only once, the DNA sequence comprising the recognition site would be single-stranded if not for the "hairpin" structure. However, in the presence of the hairpin structure, that region is effectively double stranded, thus providing a double stranded substrate for activity by restriction enzymes.

To the extent that the reaction conditions are compatible, all the primer pairs to analyze a locus or loci of interest of DNA can be mixed together for use in the method of the invention. In a preferred embodiment, all primer pairs are mixed with the template DNA in a single reaction vessel. Such a reaction vessel can be, for example, a reaction tube, or a well of a microtiter plate.

Alternatively, to avoid competition for nucleotides and to minimize primer dimers and difficulties with annealing temperatures for primers, each locus of interest or small groups of loci of interest can be amplified in separate reaction tubes or wells, and the products later pooled if desired. For example, the separate reactions can be pooled into a single reaction vessel before digestion with the restriction enzyme that generates a 5' overhang, which contains the locus of interest or SNP site, and a 3' recessed end. Preferably, the primers of each primer pair are provided in equimolar amounts. Also, especially preferably, each of the different primer pairs is provided in equimolar amounts relative to the other pairs that are being used.

In another embodiment, combinations of primer pairs that allow efficient amplification of their respective loci of interest can be used (see e.g. FIG. 2). Such

combinations can be determined prior to use in the method of the invention. Multi-well plates and PCR machines can be used to select primer pairs that work efficiently with one another. For example, gradient PCR machines, such as the Eppendorf Mastercycler® gradient PCR machine, can be used to select the optimal annealing temperature for each primer pair. Primer pairs that have similar properties can be used together in a single reaction tube.

In another embodiment, a multi-sample container including but not limited to a 96-well or more plate can be used to amplify a single locus of interest with the same primer pairs from multiple template DNA samples with optimal PCR conditions for that locus of interest. Alternatively, a separate multi-sample container can be used for amplification of each locus of interest and the products for each template DNA sample later pooled. For example, gene A from 96 different DNA samples can be amplified in microtiter plate 1, gene B from 96 different DNA samples can be amplified in microtiter plate 2, etc., and then the amplification products can be pooled.

The result of amplifying multiple loci of interest is a preparation that contains representative PCR products having the sequence of each locus of interest. For example, if DNA from only one individual is used as the template DNA and if hundreds of disease-related loci of interest were amplified from the template DNA, the amplified DNA would be a mixture of small, PCR products from each of the loci of interest. Such a preparation could be further analyzed at that time to determine the sequence at each locus of interest or at only some loci of interest. Additionally, the preparation could be stored in a manner that preserves the DNA and can be analyzed at a later time. Information contained in the amplified DNA can be revealed by any suitable method including but not limited to fluorescence detection, sequencing, gel electrophoresis, and mass spectrometry (see "Detection of Incorporated Nucleotide" section below).

II. Amplification of Loci of Interest

The template DNA can be amplified using any suitable method known in the art including but not limited to PCR (polymerase chain reaction), 3SR (self-sustained sequence reaction), LCR (ligase chain reaction), RACE-PCR (rapid amplification of cDNA ends), PLCR (a combination of polymerase chain reaction and ligase chain reaction), Q-beta phage amplification (Shah *et al., J. Medical Micro.* 33: 1435-41 (1995)), SDA (strand displacement amplification), SOE-PCR (splice overlap extension PCR), and the like. These methods can be used to design variations of the releasable

primer mediated cyclic amplification reaction explicitly described in this application. In the most preferred embodiment, the template DNA is amplified using PCR (PCR: A Practical Approach, M. J. McPherson, et al., IRL Press (1991); PCR Protocols: A Guide to Methods and Applications, Innis, et al., Academic Press (1990); and PCR Technology: Principals and Applications of DNA Amplification, H. A. Erlich, Stockton Press (1989)). PCR is also described in numerous U.S. patents, including U.S. Pat. Nos. 4,683,195; 4,683,202; 4,800,159; 4,965,188; 4,889,818; 5,075,216; 5,079,352; 5,104,792, 5,023,171; 5,091,310; and 5,066,5 84.

The components of a typical PCR reaction include but are not limited to a template DNA, primers, a reaction buffer (dependent on choice of polymerase), dNTPs (dATP, dTTP, dGTP, and dCTP) and a DNA polymerase. Suitable PCR primers can be designed and prepared as discussed above (see "Primer Design" section above). Briefly, the reaction is heated to 95°C for 2 min. to separate the strands of the template DNA, the reaction is cooled to an appropriate temperature (determined by calculating the annealing temperature of designed primers) to allow primers to anneal to the template DNA, and heated to 72°C for two minutes to allow extension.

In a preferred embodiment, the annealing temperature is increased in each of the first three cycles of amplification to reduce non-specific amplification. See also Example I, below. The TM1 of the first cycle of PCR is about the melting temperature of the 3' region of the second primer that anneals to the template DNA. The annealing temperature can be raised in cycles 2-10, preferably in cycle 2, to TM2, which is about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. If the annealing temperature is raised in cycle 2, the annealing temperature remains about the same until the next increase in annealing temperature. Finally, in any cycle subsequent to the cycle in which the annealing temperature was increased to TM2, preferably cycle 3, the annealing temperature is raised to TM3, which is about the melting temperature of the entire second primer. After the third cycle, the annealing temperature for the remaining cycles can be at about TM3 or can be further increased. In this example, the annealing temperature is increased in cycles 2 and 3. However, the annealing temperature can be increased from a low annealing temperature in cycle 1 to a high annealing temperature in cycle 2 without any further increases in temperature or the annealing temperature can progressively change from a low annealing temperature to a

high annealing temperature in any number of incremental steps. For example, the annealing temperature can be changed in cycles 2, 3, 4, 5, 6, etc.

After annealing, the temperature in each cycle is increased to an "extension" temperature to allow the primers to "extend" and then following extension the temperature in each cycle is increased to the denaturization temperature. For PCR products less than 500 base pairs in size, one can eliminate the extension step in each cycle and just have denaturization and annealing steps. A typical PCR reaction consists of 25-45 cycles of denaturation, annealing and extension as described above. However, as previously noted, one cycle of amplification (one copy) can be sufficient for practicing the invention.

In another embodiment, multiple sets of primers wherein a primer set comprises a forward primer and a reverser primer, can be used to amplify the template DNA for 1-5, 5-10, 10-15, 15-20 or more than 20 cycles, and then the amplified product is further amplified in a reaction with a single primer set or a subset of the multiple primer sets. In a preferred embodiment, a low concentration of each primer set is used to minimize primer-dimer formation. A low concentration of starting DNA can be amplified using multiple primer sets. Any number of primer sets can be used in the first amplification reaction including but not limiting to 1-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-250, 250-300, 300-350, 350-400, 400-450, 450-500, 500-1000, and greater than 1000. In another embodiment, the amplified product is amplified in a second reaction with a single primer set. In another embodiment, the amplified product is further amplified with a subset of the multiple primer pairs including but not limited to 2-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-250, and more than 250.

The multiple primer sets will amplify the loci of interest, such that a minimal amount of template DNA is not limiting for the number of loci that can be detected. For example, if template DNA is isolated from a single cell or the template DNA is obtained from a pregnant female, which comprises both maternal template DNA and fetal template DNA, low concentrations of each primer set can be used in a first amplification reaction to amplify the loci of interest. The low concentration of primers reduces the formation of primer-dimer and increases the probability that the primers will anneal to the template DNA and allow the polymerase to extend. The optimal number of cycles performed with the multiple primer sets is determined by the concentration of the primers. Following the

first amplification reaction, additional primers can be added to further amplify the loci of interest. Additional amounts of each primer set can be added and further amplified in a single reaction. Alternatively, the amplified product can be further amplified using a single primer set in each reaction or a subset of the multiple primers sets. For example, if 150 primer sets were used in the first amplification reaction, subsets of 10 primer sets can be used to further amplify the product from the first reaction.

Any DNA polymerase that catalyzes primer extension can be used including but not limited to E. coli DNA polymerase, Klenow fragment of E. coli DNA polymerase 1, T7 DNA polymerase, T4 DNA polymerase, Taq polymerase, Pfu DNA polymerase, Vent DNA polymerase, bacteriophage 29, REDTaq™ Genomic DNA polymerase, or sequenase. Preferably, a thermostable DNA polymerase is used. A "hot start" PCR can also be performed wherein the reaction is heated to 95°C for two minutes prior to addition of the polymerase or the polymerase can be kept inactive until the first heating step in cycle 1. "Hot start" PCR can be used to minimize nonspecific amplification. Any number of PCR cycles can be used to amplify the DNA, including but not limited to 2, 5, 10, 15, 20, 25, 30, 35, 40, or 45 cycles. In a most preferred embodiment, the number of PCR cycles performed is such that equimolar amounts of each loci of interest are produced.

III. Purification of Amplified DNA

Purification of the amplified DNA is not necessary for practicing the invention. However, in one embodiment, if purification is preferred, the 5' end of the primer (first or second primer) can be modified with a tag that facilitates purification of the PCR products. In a preferred embodiment, the first primer is modified with a tag that facilitates purification of the PCR products. The modification is preferably the same for all primers, although different modifications can be used if it is desired to separate the PCR products into different groups.

The tag can be a radioisotope, fluorescent reporter molecule, chemiluminescent reporter molecule, antibody, antibody fragment, hapten, biotin, derivative of biotin, photobiotin, iminobiotin, digoxigenin, avidin, enzyme, acridinium, sugar, enzyme, apoenzyme, homopolymeric oligonucleotide, hormone, ferromagnetic moiety, paramagnetic moiety, diamagnetic moiety, phosphorescent moiety, luminescent moiety, electrochemiluminescent moiety, chromatic moiety, moiety having a detectable electron

spin resonance, electrical capacitance, dielectric constant or electrical conductivity, or combinations thereof.

As one example, the 5' ends of the primers can be biotinylated (Kandpal *et al., Nucleic Acids Res. 18*:1789-1795 (1990); Kaneoka *et al., Biotechniques 10*:30-34 (1991); Green *et al., Nucleic Acids Res. 18*:6163-6164 (1990)). The biotin provides an affinity tag that can be used to purify the copied DNA from the genomic DNA or any other DNA molecules that are not of interest. Biotinylated molecules can be purified using a streptavidin coated matrix as shown in FIG. 1F, including but not limited to Streptawell, transparent, High-Bind plates from Roche Molecular Biochemicals (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog).

The PCR product of each locus of interest is placed into separate wells of a Streptavidin coated plate. Alternatively, the PCR products of the loci of interest can be pooled and placed into a streptavidin coated matrix, including but not limited to the Streptawell, transparent, High-Bind plates from Roche Molecular Biochemicals (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog).

The amplified DNA can also be separated from the template DNA using non-affinity methods known in the art, for example, by polyacrylamide gel electrophoresis using standard protocols.

IV. Digestion of Amplified DNA

The amplified DNA can be digested with a restriction enzyme that recognizes a sequence that had been provided on the first or second primer using standard protocols known within the art (FIGS. 6A-6D). Restriction enzyme digestions are performed using standard protocols well known within the art. The enzyme used depends on the restriction recognition site generated with the first or second primer. See "Primer Design" section, above, for details on restriction recognition sites generated on primers.

Type IIS restriction enzymes are extremely useful in that they cut approximately 10-20 base pairs outside of the recognition site. Preferably, the Type IIS restriction enzymes used are those that generate a 5' overhang and a recessed 3' end, including but not limited to BceA I and BsmF I (see e.g. Table 1). In a most preferred embodiment, the second primer (either forward or reverse) contains a restriction enzyme recognition sequence for BsmF I or BceA I. The Type IIS restriction enzyme BsmF I recognizes the nucleic acid sequence GGGAC, and cuts 14 nucleotides from the recognition site on the

antisense strand and 10 nucleotides from the recognition site on the sense strand. Digestion with BsmF I generates a 5' overhang of four (4) bases.

For example, if the second primer is designed so that after amplification the restriction enzyme recognition site is 13 bases from the locus of interest, then after digestion, the locus of interest is the first base in the 5' overhang (reading 3' to 5'), and the recessed 3' end is one base from the locus of interest. The 3' recessed end can be filled in with a nucleotide that is complementary to the locus of interest. One base of the overhang can be filled in using dideoxynucleotides. However, 1, 2, 3, or 4 bases of the overhang can be filled in using deoxynucleotides or a mixture of dideoxynucleotides and deoxynucleotides.

The restriction enzyme BsmF I cuts DNA ten (10) nucleotides from the recognition site on the sense strand and fourteen (14) nucleotides from the recognition site on the antisense strand. However, in a sequence dependent manner, the restriction enzyme BsmF I also cuts eleven (11) nucleotides from the recognition site on the sense strand and fifteen (15) nucleotides from the recognition site on the antisense strand. Thus, two populations of DNA molecules exist after digestion: DNA molecules cut at 10/14 and DNA molecules cut at 11/15. If the recognition site for BsmF I is 13 bases from the locus of interest in the amplified product, then DNA molecules cut at the 11/15 position will generate a 5' overhang that contains the locus of interest in the second position of the overhang (reading 3' to 5'). The 3' recessed end of the DNA molecules can be filled in with labeled nucleotides. For example, if labeled dideoxynucleotides are used, the 3' recessed end of the molecules cut at 11/15 would be filled in with one base, which corresponds to the base from the locus of interest, and the 3' recessed end of molecules cut at 10/14 would be filled in with one base, which corresponds to the locus of interest. The DNA molecules that have been cut at the 10/14 position and the DNA molecules that have been cut at the 11/15 position can be separated by size, and the incorporated nucleotides detected. This allows detection of both the nucleotide before the locus of interest, detection of the locus of interest, and potentially the three bases after the locus of interest.

Alternatively, if the base from the locus of interest and the locus of interest are different nucleotides, then the 3' recessed end of the molecules cut at 11/15 can be filled in with deoxynucleotide that is complementary to the upstream base. The remaining deoxynucleotide is washed away, and the locus of interest site can be filled in with either

labeled deoxynucleotides, unlabeled deoxynucleotides, labeled dideoxynucleotides, or unlabeled dideoxynucleotides. After the fill in reaction, the nucleotide can be detected by any suitable method. Thus, after the first fill in reaction with dNTP, the 3' recessed end of the molecules cut at 10/14 and 11/15 is from the locus of interest. The 3' recessed end can now be filled in one base, which corresponds to the locus of interest, two bases, three bases or four bases.

The restriction enzyme BceA I recognizes the nucleic acid sequence ACGGC and cuts 12 (twelve) nucleotides from the recognition site on the sense strand and 14 (fourteen) nucleotides from the recognition site on the antisense strand. If the distance from the recognition site for BceA I on the second primer is designed to be thirteen (13) bases from the locus of interest (see FIGS. 4A-4D), digestion with BceA I will generate a 5' overhang of two bases, which contains the locus of interest, and a recessed 3' end that is from the locus of interest. The locus of interest is the first nucleotide in the 5' overhang (reading 3' to 5').

Alternative cutting is also seen with the restriction enzyme BceA I, although at a much lower frequency than is seen with BsmF I. The restriction enzyme BceA I can cut thirteen (13) nucleotides from the recognition site on the sense strand and fifteen (15) nucleotides from the recognition site on the antisense strand. Thus, two populations of DNA molecules exist: DNA molecules cut at 12/14 and DNA molecules cut at 13/15. If the restriction enzyme recognition site is 13 bases from the locus of interest in the amplified product, DNA molecules cut at the 13/15 position yield a 5' overhang, which contains the locus of interest in the second position of the overhang (reading 3' to 5'). Labeled dideoxynucleotides can be used to fill in the 3' recessed end of the DNA molecules. The DNA molecules cut at 13/15 will have the base from the locus of interest filled in, and the DNA molecules cut at 12/14 will have the locus of interest site filled in. The DNA molecules cut at 13/15 and those cut at 12/14 can be separated by size, and the incorporated nucleotide detected. Thus, the alternative cutting can be used to obtain additional sequence information.

Alternatively, if the two bases in the 5' overhang are different, the 3' recessed end of the DNA molecules, which were cut at 13/15, can be filled in with the deoxynucleotide complementary to the first base in the overhang, and excess deoxynucleotide washed away. After filling in, the 3' recessed end of the DNA molecules that were cut at 12/14 and the DNA molecules that were cut at 13/15 are from

the locus of interest. The 3' recessed ends can be filled with either labeled dideoxynucleotides, unlabeled dideoxynucleotides, labeled deoxynucleotides, or unlabeled deoxynucleotides.

If the primers provide different restriction sites for certain of the loci of interest that were copied, all the necessary restriction enzymes can be added together to digest the copied DNA simultaneously. Alternatively, the different restriction digests can be made in sequence, for example, using one restriction enzyme at a time, so that only the product that is specific for that restriction enzyme is digested.

Optimal restriction enzyme digestion conditions, including but not limited to the concentration of enzyme, temperature, buffer conditions, and the time of digestion can be optimized for each restriction enzyme. For example, the alternative cutting seen with the type IIS restriction enzyme BsmF I can be reduced, if desired, by performing the restriction enzyme digestion at lower temperatures including but not limited to 25-16°, 16-12°C, 12-8°C, 8-4°C, or 4-0°C.

V. Incorporation of Labeled Nucleotides

Digestion with the restriction enzyme that recognizes the sequence on the second primer generates a recessed 3' end and a 5' overhang, which contains the locus of interest (FIG. 1G). The recessed 3' end can be filled in using the 5' overhang as a template in the presence of unlabeled or labeled nucleotides or a combination of both unlabeled and labeled nucleotides. The nucleotides can be labeled with any type of chemical group or moiety that allows for detection including but not limited to radioactive molecules, fluorescent molecules, antibodies, antibody fragments, haptens, carbohydrates, biotin, derivatives of biotin, phosphorescent moieties, luminescent moieties, electrochemiluminescent moieties, chromatic moieties, and moieties having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. The nucleotides can be labeled with one or more than one type of chemical group or moiety. Each nucleotide can be labeled with the same chemical group or moiety. Alternatively, each different nucleotide can be labeled with a different chemical group or moiety. The labeled nucleotides can be dNTPs, ddNTPs, or a mixture of both dNTPs and ddNTPs. The unlabeled nucleotides can be dNTPs, ddNTPs or a mixture of both dNTPs and ddNTPs.

Any combination of nucleotides can be used to incorporate nucleotides including but not limited to unlabeled deoxynucleotides, labeled deoxynucleotides, unlabeled

dideoxynucleotides, labeled dideoxynucleotides, a mixture of labeled and unlabeled deoxynucleotides, a mixture of labeled and unlabeled dideoxynucleotides, a mixture of labeled deoxynucleotides and labeled dideoxynucleotides, a mixture of labeled deoxynucleotides and unlabeled dideoxynucleotides, a mixture of unlabeled deoxynucleotides and unlabeled dideoxynucleotides, a mixture of unlabeled deoxynucleotides and labeled dideoxynucleotides, dideoxynucleotide analogues, deoxynucleotide analogues, a mixture of dideoxynucleotide analogues and deoxynucleotide analogues, phosphorylated nucleoside analogues, 2'-deoxynucleotide-5'-triphosphate, and modified 2'-deoxynucleotide-5'-triphosphate.

For example, as shown in FIG. 1H, in the presence of a polymerase, the 3' recessed end can be filled in with fluorescent ddNTP using the 5' overhang as a template. The incorporated ddNTP can be detected using any suitable method including but not limited to fluorescence detection.

All four nucleotides can be labeled with different fluorescent groups, which will allow one reaction to be performed in the presence of all four labeled nucleotides. Alternatively, four separate "fill in" reactions can be performed for each locus of interest; each of the four reactions will contain a different labeled nucleotide (e.g. ddATP*, ddTTP*, ddGTP*, or ddCTP*, where * indicates a labeled nucleotide). Each nucleotide can be labeled with different chemical groups or the same chemical groups. The labeled nucleotides can be dideoxynucleotides or deoxynucleotides.

In another embodiment, nucleotides can be labeled with fluorescent dyes including but not limited to fluorescein, pyrene, 7-methoxycoumarin, Cascade Blue.TM., Alexa Flur 350, Alexa Flur 430, Alexa Flur 488, Alexa Flur 532, Alexa Flur 546, Alexa Flur 568, Alexa Flur 594, Alexa Flur 633, Alexa Flur 647, Alexa Flur 660, Alexa Flur 680, AMCA-X, dialkylaminocoumarin, Pacific Blue, Marina Blue, BODIPY 493/503, BODIPY Fl-X, DTAF, Oregon Green 500, Dansyl-X, 6-FAM, Oregon Green 488, Oregon Green 514, Rhodamine Green-X, Rhodol Green, Calcein, Eosin, ethidium bromide, NBD, TET, 2', 4', 5', 7' tetrabromosulfonefluorescien, BODIPY-R6G, BODIPY-Fl BR2, BODIPY 530/550, HEX, BODIPY 558/568, BODIPY-TMR-X., PyMPO, BODIPY 564/570, TAMRA, BODIPY 576/589, Cy3, Rhodamine Red-x, BODIPY 581/591, carboxyXrhodamine, Texas Red-X, BODIPY-TR-X., Cy5, SpectrumAqua, SpectrumGreen #1, SpectrumGreen #2, SpectrumOrange, SpectrumRed, or naphthofluorescein.

In another embodiment, the "fill in" reaction can be performed with fluorescently labeled dNTPs, wherein the nucleotides are labeled with different fluorescent groups. The incorporated nucleotides can be detected by any suitable method including but not limited to Fluorescence Resonance Energy Transfer (FRET).

In another embodiment, a mixture of both labeled ddNTPs and unlabeled dNTPs can be used for filling in the recessed 3' end of the SNP or locus of interest. Preferably, the 5' overhang consists of more than one base, including but not limited to 2, 3, 4, 5, 6 or more than 6 bases. For example, if the 5' overhang consists of the sequence "XGAA," wherein X is the locus of interest, e.g. SNP, then filling in with a mixture of labeled ddNTPs and unlabeled dNTPs will produce several different DNA fragments. If a labeled ddNTP is incorporated at position "X," the reaction will terminate and a single labeled base will be incorporated. If however, an unlabeled dNTP is incorporated, the polymerase continues to incorporate other bases until a labeled ddNTP is incorporated. If the first two nucleotides incorporated are dNTPs, and the third is a ddNTP, the 3' recessed end will be extend by three bases. This DNA fragment can be separated from the other DNA fragments that were extended by 1, 2, or 4 bases by size. A mixture of labeled ddNTPs and unlabeled dNTPs will allow all bases of the overhang to be filled in, and provides additional sequence information about the locus of interest, e.g. SNP (see FIGS. 7E and 9D).

After incorporation of the labeled nucleotide, the amplified DNA can be digested with a restriction enzyme that recognizes the sequence provided by the first primer. For example, in FIG 11, the amplified DNA is digested with a restriction enzyme that binds to region "a," which releases the DNA fragment containing the incorporated nucleotide from the streptavidin matrix.

Alternatively, one primer of each primer pair for each locus of interest can be attached to a solid support matrix including but not limited to a well of a microtiter plate. For example, streptavidin-coated microtiter plates can be used for the amplification reaction with a primer pair, wherein one primer is biotinylated. First, biotinylated primers are bound to the streptavidin-coated microtiter plates. Then, the plates are used as the reaction vessel for PCR amplification of the loci of interest. After the amplification reaction is complete, the excess primers, salts, and template DNA can be removed by washing. The amplified DNA remains attached to the microtiter plate. The amplified DNA can be digested with a restriction enzyme that recognizes a sequence on

the second primer and generates a 5' overhang, which contains the locus of interest. The digested fragments can be removed by washing. After digestion, the SNP site or locus of interest is exposed in the 5' overhang. The recessed 3' end is filled in with a labeled nucleotide, including but not limited to, fluorescent ddNTP in the presence of a polymerase. The labeled DNA can be released into the supernatant in the microtiter plate by digesting with a restriction enzyme that recognizes a sequence in the 5' region of the first primer.

In another embodiment, one nucleotide can be used to determine the sequence of multiple alleles of a gene. A nucleotide that terminates the elongation reaction can be used to determine the sequence of multiple alleles of a gene. At one allele, the terminating nucleotide is complementary to the locus of interest in the 5' overhang of said allele. The nucleotide is incorporated and terminates the reaction. At a different allele, the terminating nucleotide is not complementary to the locus of interest, which allows a non-terminating nucleotide to be incorporated at the locus of interest of the different allele. However, the terminating nucleotide is complementary to a nucleotide downstream from the locus of interest in the 5' overhang of said different allele. The sequence of the alleles can be determined by analyzing the patterns of incorporation of the terminating nucleotide. The terminating nucleotide can be labeled or unlabeled.

In a another embodiment, the terminating nucleotide is a nucleotide that terminates or hinders the elongation reaction including but not limited to a dideoxynucleotide, a dideoxynucleotide derivative, a dideoxynucleotide analog, a dideoxynucleotide homolog, a dideoxynucleotide with a sulfur chemical group, a deoxynucleotide, a deoxynucleotide derivative, a deoxynucleotide homolog, a deoxynucleotide analog, a deoxynucleotide with a sulfur chemical group, arabinoside triphosphate, a arabinoside triphosphate analog, a arabinoside triphosphate homolog, or an arabinoside derivative.

In another embodiment, a terminating nucleotide labeled with one signal generating moiety tag, including but not limited to a fluorescent dye, can be used to determine the sequence of the alleles of a locus of interest. The use of a single nucleotide labeled with one signal generating moiety tag eliminates any difficulties that can arise when using different fluorescent moieties. In addition, using one nucleotide labeled with one signal generating moiety tag to determine the sequence of alleles of a locus of interest reduces the number of reactions, and eliminates pipetting errors.

For example, if the second primer contains the restriction enzyme recognition site for BsmFI, digestion will generate a 5' overhang of 4 bases. The second primer can be designed such that the locus of interest is located in the first position of the overhang. A representative overhang is depicted below, where R represents the locus of interest:

	5' CAC				
	3' GTG	R	T	G	G
Overhang position		1	2	3	4

One nucleotide with one signal generating moiety tag can be used to determine whether the variable site is homozygous or heterozygous. For example, if the variable site is adenine (A) or guanine (G), then either adenine or guanine can be used to determine the sequence of the alleles of the locus of interest, provided that there is an adenine or guanine in the overhang at position 2, 3, or 4.

For example, if the nucleotide in position 2 of the overhang is thymidine, which is complementary to adenine, then labeled ddATP, unlabeled dCTP, dGTP, and dTTP can be used to determine the sequence of the alleles of the locus of interest. The ddATP can be labeled with any signal generating moiety including but not limited to a fluorescent dye. If the template DNA is homozygous for adenine, then labeled ddATP* will be incorporated at position 1 complementary to the overhang at the alleles, and no nucleotide incorporation will be seen at position 2, 3 or 4 complementary to the overhang.

Allele 1	5' CCC	A*			
	3' GGG	T	T	G	G
Overhang position		1	2	3	4

Allele 2	5' CCC	A*			
	3' GGG	T	T	G	G
Overhang position		1	2	3	4

One signal will be seen corresponding to incorporation of labeled ddATP at position 1 complementary to the overhang, which indicates that the individual is homozygous for adenine at this position. This method of labeling eliminates any

difficulties that may arise from using different dyes that have different quantum coefficients.

Homozygous guanine:

If the template DNA is homozygous for guanine, then no ddATP will be incorporated at position 1 complementary to the overhang, but ddATP will be incorporated at the first available position, which in this case is position 2 complementary to the overhang. For example, if the second position in the overhang corresponds to a thymidine, then:

Allele 1	5' CCC	**G**	**A***		
	3' GGG	C	T	G	G
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	**A***		
	3' GGG	C	T	G	G
Overhang position		1	2	3	4

One signal will be seen corresponding to incorporation of ddATP at position 2 complementary to the overhang, which indicates that the individual is homozygous for guanine. The molecules that are filled in at position 2 complementary to the overhang will have a different molecular weight than the molecules filled in at position 1 complementary to the overhang.

Heterozygous condition:

Allele 1	5' CCC	**A***			
	3' GGG	T	T	G	G
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	**A***		
	3' GGG	C	T	G	G
Overhang position		1	2	3	4

Two signals will be seen; the first signal corresponds to the ddATP filled in at position one complementary to the overhang and the second signal corresponds to the ddATP filled in at position 2 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by a single base pair, which allows easy detection and quantitation of the signals. Molecules filled in at position one can be distinguished from molecules filled in at position two using any method that discriminates based on molecular weight including but not limited to gel electrophoresis, capillary gel electrophoresis, DNA sequencing, and mass spectrometry. It is not necessary that the nucleotide be labeled with a chemical moiety; the DNA molecules corresponding to the different alleles can be separated based on molecular weight.

If position 2 of the overhang is not complementary to adenine, it is possible that positions 3 or 4 may be complementary to adenine. For example, position 3 of the overhang may be complementary to the nucleotide adenine, in which case labeled ddATP may be used to determine the sequence of both alleles.

Homozygous for adenine:

Allele 1	5' CCC	A*			
	3' GGG	T	G	T	G
Overhang position		1	2	3	4
Allele 2	5' CCC	A*			
	3' GGG	T	G	T	G
Overhang position		1	2	3	4

Homozygous for guanine:

Allele 1	5' CCC	G	C	A*	
	3' GGG	C	G	T	G
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	C	**A***	
	3' GGG	C	G	T	G
Overhang position		1	2	3	4

Heterozygous:

Allele 1	5' CCC	**A***			
	3' GGG	T	G	T	G
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	C	**A***	
	3' GGG	C	G	T	G
Overhang position		1	2	3	4

Two signals will be seen; the first signal corresponds to the ddATP filled in at position 1 complementary to the overhang and the second signal corresponds to the ddATP filled in at position 3 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by two bases, which can be detected using any method that discriminates based on molecular weight.

Alternatively, if positions 2 and 3 are not complementary to adenine (*i.e* positions 2 and 3 of the overhang correspond to guanine, cytosine, or adenine) but position 4 is complementary to adenine, labeled ddATP can be used to determine the sequence of both alleles.

Homozygous for adenine:

Allele 1	5' CCC	**A***			
	3' GGG	T	G	G	T
Overhang position		1	2	3	4

Allele 2	5' CCC	**A***			
	3' GGG	T	G	G	T
Overhang position		1	2	3	4

One signal will be seen that corresponds to the molecular weight of molecules filled in with ddATP at position one complementary to the overhang, which indicates that the individual is homozygous for adenine at the variable site.

Homozygous for guanine:

Allele 1	5' CCC	**G**	C	C	**A***
	3' GGG	C	G	G	T
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	C	C	**A***
	3' GGG	C	G	G	T
Overhang position		1	2	3	4

One signal will be seen that corresponds to the molecular weight of molecules filled in at position 4 complementary to the overhang, which indicates that the individual is homozygous for guanine.

Heterozygous:

Allele 1	5' CCC	**A***			
	3' GGG	**T**	G	G	T
Overhang position		1	2	3	4

Allele 2	5' CCC	**G**	C	C	**A***
	3' GGG	**C**	G	G	T
Overhang position		1	2	3	4

Two signals will be seen; the first signal corresponds to the ddATP filled in at position one complementary to the overhang and the second signal corresponds to the ddATP filled in at position 4 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by three

bases, which allows detection and quantitation of the signals. The molecules filled in at position 1 and those filled in at position 4 can be distinguished based on molecular weight.

As discussed above, if the variable site contains either adenine or guanine, either labeled adenine or labeled guanine can be used to determine the sequence of both alleles. If positions 2, 3, or 4 of the overhang are not complementary to adenine but one of the positions is complementary to a guanine, then labeled ddGTP can be used to determine whether the template DNA is homozygous or heterozygous for adenine or guanine. For example, if position 3 in the overhang corresponds to a cytosine then the following signals will be expected if the template DNA is homozygous for guanine, homozygous for adenine, or heterozygous:

Homozygous for guanine:

Allele 1	5' CCC	**G***			
	3' GGG	C	T	C	T
Overhang position		1	2	3	4
Allele 2	5' CCC	**G***			
	3' GGG	C	T	C	T
Overhang position		1	2	3	4

One signal will be seen that corresponds to the molecular weight of molecules filled in with ddGTP at position one complementary to the overhang, which indicates that the individual is homozygous for guanine.

Homozygous for adenine:

Allele 1	5' CCC	**A**	**A**	**G***	
	3' GGG	T	T	C	T
Overhang position		1	2	3	4
Allele 2	5' CCC	**A**	**A**	**G***	

	3' GGG	T	T	C	T
Overhang position		1	2	3	4

One signal will be seen that corresponds to the molecular weight of molecules filled in at position 3 complementary to the overhang, which indicates that the individual is homozygous for adenine at the variable site.

Heterozygous:

Allele 1	5' CCC	**G***			
	3' GGG	C	T	C	T
Overhang position		1	2	3	4
Allele 2	5' CCC	**A**	**A**	**G***	
	3' GGG	T	T	C	T
Overhang position		1	2	3	4

Two signals will be seen; the first signal corresponds to the ddGTP filled in at position one complementary to the overhang and the second signal corresponds to the ddGTP filled in at position 3 complementary to the overhang. The two signals can be separated based on molecular weight; allele 1 and allele 2 will be separated by two bases, which allows easy detection and quantitation of the signals.

Some type IIS restriction enzymes also display alternative cutting as discussed above. For example, BsmFI will cut at 10/14 and 11/15 from the recognition site. However, the cutting patterns are not mutually exclusive; if the 11/15 cutting pattern is seen at a particular sequence, 10/14 cutting is also seen. If the restriction enzyme BsmF I cuts at 10/14 from the recognition site, the 5' overhang will be $X_1X_2X_3X_4$. If BsmF I cuts 11/15 from the recognition site, the 5' overhang will be $X_0X_1X_2X_3$. If position X_0 of the overhang is complementary to the labeled nucleotide, the labeled nucleotide will be incorporated at position X_0 and provides an additional level of quality assurance. It provides additional sequence information.

For example, if the variable site is adenine or guanine, and position 3 in the overhang is complementary to adenine, labeled ddATP can be used to determine the

genotype at the variable site. If position 0 of the 11/15 overhang contains the nucleotide complementary to adenine, ddATP will be filled in and an additional signal will be seen.

Heterozygous:

10/14 Allele 1	5' CCA	**A***			
	3' GGT	T	G	T	G
Overhang position		1	2	3	4
10/14 Allele 2	5' CCA	**G**	C	**A***	
	3' GGT	C	G	T	G
Overhang position		1	2	3	4
11/15 Allele 1	5' CC	**A***			
	3' GG	T	T	G	T
Overhang position		0	1	2	3
11/15 Allele 2	5' CC	**A***			
	3' GG	T	C	G	T
Overhang position		0	1	2	3

Three signals are seen; one corresponding to the ddATP incorporated at position 0 complementary to the overhang, one corresponding to the ddATP incorporated at position 1 complementary to the overhang, and one corresponding to the ddATP incorporated at position 3 complementary to the overhang. The molecules filled in at position 0, 1, and 3 complementary to the overhang differ in molecular weight and can be separated using any technique that discriminates based on molecular weight including but not limited to gel electrophoresis, and mass spectrometry.

For quantitating the ratio of one allele to another allele or when determining the relative amount of a mutant DNA sequence in the presence of wild type DNA sequence, an accurate and highly sensitive method of detection must be used. The alternate cutting displayed by type IIS restriction enzymes may increase the difficulty of determining ratios of one allele to another allele because the restriction enzyme may not display the

alternate cutting (11/15) pattern on the two alleles equally. For example, allele 1 may be cut at 10/14 80% of the time, and 11/15 20% of the time. However, because the two alleles may differ in sequence, allele 2 may be cut at 10/14 90% of the time, and 11/15 20% of the time.

For purposes of quantitation, the alternate cutting problem can be eliminated when the nucleotide at position 0 of the overhang is not complementary to the labeled nucleotide. For example, if the variable site corresponds to adenine or guanine, and position 3 of the overhang is complementary to adenine (*i.e*, a thymidine is located at position 3 of the overhang), labeled ddATP can be used to determine the genotype of the variable site. If position 0 of the overhang generated by the 11/15 cutting properties is not complementary to adenine, (*i.e*, position 0 of the overhang corresponds to guanine, cytosine, or adenine) no additional signal will be seen from the fragments that were cut 11/15 from the recognition site. Position 0 complementary to the overhang can be filled in with unlabeled nucleotide, eliminating any complexity seen from the alternate cutting pattern of restriction enzymes. This method provides a highly accurate method for quantitating the ratio of a variable site including but not limited to a mutation, or a single nucleotide polymorphism.

For instance, if SNP X can be adenine or guanine, this method of labeling allows quantitation of the alleles that correspond to adenine and the alleles that correspond to guanine, without determining if the restriction enzyme displays any differences between the alleles with regard to alternate cutting patterns.

Heterozygous:

10/14 Allele 1	5' CCG	**A***			
	3' GGC	T	G	T	G
Overhang position		1	2	3	4
10/14 Allele 2	5' CCG	**G**	**C**	**A***	
	3' GGC	C	G	T	G
Overhang position		1	2	3	4

The overhang generated by the alternate cutting properties of BsmF I is depicted below:

	11/15 Allele 1	5' CC				
		3' GG	C	T	G	T
	Overhang position		0	1	2	3
	11/15 Allele 2	5' CC				
		3' GG	C	C	G	T
	Overhang position		0	1	2	3

After filling in with labeled ddATP and unlabeled dGTP, dCTP, dTTP, the following molecules would be generated:

	11/15 Allele 1	5' CC	G	**A***		
		3' GG	C	T	G	T
	Overhang position		0	1	2	3
	11/15 Allele 2	5' CC	G	**G**	C	**A***
		3' GG	C	C	G	T
	Overhang position		0	1	2	3

Two signals are seen; one corresponding to the molecules filled in with ddATP at position one complementary to the overhang and one corresponding to the molecules filled in with ddATP at position 3 complementary to the overhang. Position 0 of the 11/15 overhang is filled in with unlabeled nucleotide, which eliminates any difficulty in quantitating a ratio for the nucleotide at the variable site on allele 1 and the nucleotide at the variable site on allele 2.

Any nucleotide can be used including adenine, adenine derivatives, adenine homologues, guanine, guanine derivatives, guanine homologues, cytosine, cytosine derivatives, cytosine homologues, thymidine, thymidine derivatives, or thymidine homologues, or any combinations of adenine, adenine derivatives, adenine homologues, guanine, guanine derivatives, guanine homologues, cytosine, cytosine derivatives, cytosine homologues, thymidine, thymidine derivatives, or thymidine homologues.

The nucleotide can be labeled with any chemical group or moiety, including but not limited to radioactive molecules, fluorescent molecules, antibodies, antibody fragments, haptens, carbohydrates, biotin, derivatives of biotin, phosphorescent moieties, luminescent moieties, electrochemiluminescent moieties, chromatic moieties, and moieties having a detectable electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. The nucleotide can be labeled with one or more than one type of chemical group or moiety.

In another embodiment, labeled and unlabeled nucleotides can be used. Any combination of deoxynucleotides and dideoxynucleotides can be used including but not limited to labeled dideoxynucleotides and labeled deoxynucleotides; labeled dideoxynucleotides and unlabeled deoxynucleotides; unlabeled dideoxynucleotides and unlabeled deoxynucleotides; and unlabeled dideoxynucleotides and labeled deoxynucleotides.

In another embodiment, nucleotides labeled with a chemical moiety can be used in the PCR reaction. Unlabeled nucleotides then are used to fill-in the 5' overhangs generated after digestion with the restriction enzyme. An unlabeled terminating nucleotide can be used to in the presence of unlabeled nucleotides to determine the sequence of the alleles of a locus of interest.

For example, if labeled dTTP was used in the PCR reaction, the following 5' overhang would be generated after digestion with BsmF I:

10/14 Allele 1	5' CT*G	**A**			
	3' GA C	T	G	T	G
Overhang position		1	2	3	4
10/14 Allele 2	5' CT*G	**G**	C	**A**	
	3' GA C	C	G	T	G
Overhang position		1	2	3	4

Unlabeled ddATP, unlabeled dCTP, unlabeled dGTP, and unlabeled dTTP can be used to fill-in the 5' overhang. Two signals will be generated; one signal corresponds to the DNA molecules filled in with unlabeled ddATP at position 1 complementary to the overhang and the second signal corresponds to DNA molecules filled in with unlabeled

ddATP at position 3 complementary to the overhang. The DNA molecules can be separated based on molecular weight and can be detected by the fluorescence of the dTTP, which was incorporated during the PCR reaction.

The labeled DNA loci of interest sites can be analyzed by a variety of methods including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, and other methods of sequencing, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry or by DNA hybridization techniques including Southern Blots, Slot Blots, Dot Blots, and DNA microarrays, wherein DNA fragments would be useful as both "probes" and "targets," ELISA, fluorimetry, and Fluorescence Resonance Energy Transfer (FRET).

This method of labeling is extremely sensitive and allows the detection of alleles of a locus of interest that are in various ratios including but not limited to 1:1, 1:2, 1:3, 1:4, 1:5, 1:6-1:10, 1:11-1:20, 1:21-1:30, 1:31-1:40, 1:41-1:50, 1:51-1:60, 1:61-1:70, 1:71-1:80, 1:81-1:90, 1:91:1:100, 1:101-1:200, 1:250, 1:251-1:300, 1:301-1:400, 1:401-1:500, 1:501-1:600, 1:601-1:700, 1:701-1:800, 1:801-1:900, 1:901-1:1000, 1:1001-1:2000, 1:2001-1:3000, 1:3001-1:4000, 1:4001-1:5000, 1:5001-1:6000, 1:6001-1:7000, 1:7001-1:8000, 1:8001-1:9000, 1:9001-1:10,000; 1:10,001-1:20,000, 1:20,001:1:30,000, 1:30,001-1:40,000, 1:40,001-1:50,000, and greater than 1:50,000.

For example, this method of labeling allows one nucleotide labeled with one signal generating moiety to be used to determine the sequence of alleles at a SNP locus, or detect a mutant allele amongst a population of normal alleles, or detect an allele encoding antibiotic resistance from a bacterial cell amongst alleles from antibiotic sensitive bacteria, or detect an allele from a drug resistant virus amongst alleles from drug-sensitive virus, or detect an allele from a non-pathogenic bacterial strain amongst alleles from a pathogenic bacterial strain.

As shown above, a single nucleotide can be used to determine the sequence of the alleles at a particular locus of interest. This method is especially useful for determining if an individual is homozygous or heterozygous for a particular mutation or to determine the sequence of the alleles at a particular SNP site. This method of labeling eliminates any errors caused by the quantum coefficients of various dyes. It also allows the reaction to

proceed in a single reaction vessel including but not limited to a well of a microtiter plate, or a single eppendorf tube.

This method of labeling is especially useful for the detection of multiple genetic signals in the same sample. For example, this method is useful for the detection of fetal DNA in the blood, serum, or plasma of a pregnant female, which contains both maternal DNA and fetal DNA. The maternal DNA and fetal DNA may be present in the blood, serum or plasma at ratios such as 97:3; however, the above-described method can be used to detect the fetal DNA. This method of labeling can be used to detect two, three, or four different genetic signals in the sample population

This method of labeling is especially useful for the detection of a mutant allele that is among a large population of wild type alleles. Furthermore, this method of labeling allows the detection of a single mutant cell in a large population of wild type cells. For example, this method of labeling can be used to detect a single cancerous cell among a large population of normal cells. Typically, cancerous cells have mutations in the DNA sequence. The mutant DNA sequence can be identified even if there is a large background of wild type DNA sequence. This method of labeling can be used to screen, detect, or diagnosis any type of cancer including but not limited to colon, renal, breast, bladder, liver, kidney, brain, lung, prostate, and cancers of the blood including leukemia.

This labeling method can also be used to detect pathogenic organisms, including but not limited to bacteria, fungi, viruses, protozoa, and mycobacteria. It can also be used to discriminate between pathogenic strains of microorganism and non-pathogenic strains of microorganisms including but not limited to bacteria, fungi, viruses, protozoa, and mycobacteria.

For example, there are several strains of *Escherichia coli* (*E. coli*), and most are non-pathogenic. However, several strains, such as *E. coli O157* are pathogenic. There are genetic differences between non-pathogenic *E. coli* strains and pathogenic *E. coli*. The above described method of labeling can be used to detect pathogenic microorganisms in a large population of non-pathogenic organisms, which are sometimes associated with the normal flora of an individual.

VI. Analysis of the locus of interest

The loci of interest can be analyzed by a variety of methods including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an

automated DNA sequencing machine, (e.g. the ABI Prism 3100 Genetic Analyzer or the ABI Prism 3700 Genetic Analyzer), microchannel electrophoresis, and other methods of sequencing, Sanger dideoxy sequencing, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry or by DNA hybridization techniques including Southern Blot, Slot Blot, Dot Blot, and DNA microarray, wherein DNA fragments would be useful as both "probes" and "targets," ELISA, fluorimetry, fluorescence polarization, and Fluorescence Resonance Energy Transfer (FRET).

The loci of interest can be analyzed using gel electrophoresis followed by fluorescence detection of the incorporated nucleotide. Another method to analyze or read the loci of interest is to use a fluorescent plate reader or fluorimeter directly on the 96-well streptavidin coated plates. The plate can be placed onto a fluorescent plate reader or scanner such as the Pharmacia 9200 Typhoon to read each locus of interest.

Alternatively, the PCR products of the loci of interest can be pooled and after "filling in" (FIG. 10), the products can be separated by size, using any method appropriate for the same, and then analyzed using a variety of techniques including but not limited to fluorescence detection, DNA sequencing gel, capillary electrophoresis on an automated DNA sequencing machine, microchannel electrophoresis, other methods of sequencing, Sanger dideoxy sequencing, DNA hybridization techniques including Southern Blot, Slot Blot, Dot Blot, and DNA microarray, mass spectrometry, time of flight mass spectrometry, quadrupole mass spectrometry, magnetic sector mass spectrometry, electric sector mass spectrometry infrared spectrometry, ultraviolet spectrometry, palentiostatic amperometry. For example, polyacrylamide gel electrophoresis can be used to separate DNA by size and the gel can be scanned to determine the color of fluorescence in each band (using e.g., ABI 377 DNA sequencing machine or a Pharmacia Typhoon 9200).

In another embodiment, the sequence of the locus of interest can be determined by detecting the incorporation of a nucleotide that is 3' to the locus of interest, wherein said nucleotide is a different nucleotide from the possible nucleotides at the locus of interest. This embodiment is especially useful for the sequencing and detection of SNPs. The efficiency and rate at which DNA polymerases incorporate nucleotides varies for each nucleotide.

According to the data from the Human Genome Project, 99% of all SNPs are binary. The sequence of the human genome can be used to determine a nucleotide that is 3' to the SNP of interest. When a nucleotide that is 3' to the SNP site differs from the possible nucleotides at the SNP site, a nucleotide that is one or more than one base 3' to the SNP can be used to determine the sequence of the SNP site.

For example, suppose the sequence of SNP X on chromosome 13 is to be determined. The sequence of the human genome indicates that SNP X can either be adenosine or guanine and that a nucleotide 3' to the locus of interest is a thymidine. A primer that contains a restriction enzyme recognition site for BsmF I, which is designed to be 13 bases from the locus of interest after amplification, is used to amplify a DNA fragment containing SNP X. Digestion with the restriction enzyme BsmF I generates a 5' overhang that contains the locus of interest, which can either be adenosine or guanine. The digestion products can be split into two "fill in" reactions: one contains dTTP, and the other reaction contains dCTP. If the locus of interest is homozygous for guanine, only the DNA molecules that were mixed with dCTP will be filled in. If the locus of interest is homozygous for adenosine, only the DNA molecules that were mixed with dTTP will be filled in. If the locus of interest is heterozygous, the DNA molecules that were mixed with dCTP will be filled in as well as the DNA molecules that were mixed with dTTP. After washing to remove the excess dNTP, the samples are filled in with labeled ddATP, which is complimentary to the nucleotide (thymidine) that is 3' to the locus of interest. The DNA molecules that were filled in by the previous reaction will be filled in with labeled ddATP. If the individual is homozygous for adenosine, the DNA molecules that were mixed with dTTP subsequently will be filled in with the labeled ddATP. However, the DNA molecules that were mixed with dCTP, would not have incorporated that nucleotide, and therefore, could not incorporate the ddATP. Detection of labeled ddATP only in the molecules that were mixed with dTTP indicates that the nucleotide at SNP X on chromosome 13 is adenosine.

In another embodiment, large scale screening for the presence or absence of single nucleotide polymorphisms or mutations can be performed. One to tens to hundreds to thousands of loci of interest on a single chromosome or on multiple chromosomes can be amplified with primers as described above in the "Primer Design" section. The primers can be designed so that each amplified loci of interest is of a different size (FIG. 2). The multiple loci of interest can be of a DNA sample from one individual

representing multiple loci of interest on a single chromosome, multiple chromosomes, multiple genes, a single gene, or any combination thereof.

When human data is being analyzed, the known sequence can be a specific sequence that has been determined from one individual (including e.g. the individual whose DNA is currently being analyzed), or it can be a consensus sequence such as that published as part of the human genome.

Ratio of Alleles at Heterozygous Locus of Interest

In one embodiment, the ratio of alleles at a heterozygous locus of interest can be calculated. The intensity of a nucleotide at the loci of interest can be quantified using any number of computer programs including but not limited to GeneScan and ImageQuant. For example, for a heterozygous SNP, there are two nucleotides , and each may should be present in a 1:1 ratio. In a preferred embodiment, the ratio of multiple heterozygous SNPs can be calculated.

In one embodiment, the ratio for a variable nucleotide at alleles at a heterozygous locus of interest can be calculated. The intensity of a each variable nucleotide present at the loci of interest can be quantified using any number of computer programs including but not limited to GeneScan and ImageQuant. For example, for a heterozygous SNP, there are will be two nucleotides present, and each may be present in a 1:1 ratio. In a preferred embodiment, the ratio of multiple heterozygous SNPs can be calculated.

In another embodiment, the ratio of alleles at a heterozygous locus of interest on a chromosome is summed and compared to the ratio of alleles at a heterozygous locus of interest on a different chromosome. In a preferred embodiment, the ratio of alleles at multiple heterozygous loci of interest on a chromosome is summed and compared to the ratio of alleles at multiple heterozygous loci of interest on a different chromosome. The ratio obtained from SNP 1, SNP 2, SNP 3, SNP 4, etc on chromosome 1 can be summed. This ratio can then be compared to the ratio obtained from SNP A, SNP B, SNP C, SNP D, etc.

For example, 100 SNPs can be analyzed on chromosome 1. Of these 100 SNPs, assume 50 are heterozygous. The ratio of the alleles at heterozygous SNPs on chromosome 1 can be summed, and should give a ratio of approximately 50:50. Likewise, of 100 SNPs analyzed on chromosome 21, assume 50 are heterozygous. The ratio of alleles at heterozygous SNPs on chromosome 21 is summed. With a normal number of chromosomes, the ratio should be approximately 50:50, and thus there should

be no difference between the ratio obtained from chromosome 1 and 21. However, if there is an additional copy of chromosome 21, an additional allele will be provided, and the ratio should be approximately 66:33. Thus, the ratio for nucleotides at heterozygous SNPs can be used to detect the presence or absence of chromosomal abnormalities. Any chromosomal abnormality can be detected including aneuploidy, polyploidy, inversion, a trisomy, a monosomy, duplication, deletion, deletion of a part of a chromosome, addition, addition of a part of chromosome, insertion, a fragment of a chromosome, a region of a chromosome, chromosomal rearrangement, and translocation. The method is especially useful for the detection of trisomy 13, trisomy 18, trisomy 21, XXY, and XYY.

The present invention provides a method to quantitate a ratio for the alleles at a heterozygous locus of interest. The loci of interest include but are not limited to single nucleotide polymorphisms, mutations. There is no need to amplify the entire sequence of a gene or to quantitate the amount of a particular gene product. The present invention does not rely on quantitative PCR.

Detection of Fetal Chromosomal Abnormalities

As discussed above in the section entitled "DNA template," the template DNA can be obtained from a sample of a pregnant female, wherein the template DNA comprises maternal template DNA and fetal template DNA. In one embodiment, the template DNA is obtained from the blood of a pregnant female. In a preferred embodiment, the template DNA is obtained from the plasma or serum from the blood of a pregnant female.

In one embodiment, the template DNA from the sample from the pregnant female comprises both maternal template DNA and fetal template DNA. In another embodiment, maternal template DNA is obtained from any nucleic acid containing source including but not limited to cell, tissue, blood, serum, plasma, saliva, urine, tears, vaginal secretion, lymph fluid, cerebrospinal fluid, mucosa secretion, peritoneal fluid, ascitic fluid, fecal matter, or body exudates, and sequenced to identify homozygous or heterozygous loci of interest, which are the loci of interest analyzed on the template DNA obtained from the sample from the pregnant female.

In a preferred embodiment, the sequence of the alleles of multiple loci of interest on maternal template DNA is determined to identify homozygous loci of interest. In another embodiment, the sequence of the alleles of multiple loci of interest on maternal template DNA is determined to identify heterozygous loci of interest. The sequence of

the alleles of multiple loci of interest on maternal template DNA can be determined in a single reaction or in multiple reactions.

For example, if 100 maternal loci of interest on chromosome 21 and 100 maternal loci of interest on chromosome 1 are analyzed, one would predict approximately 50 loci of interest on each chromosome to be homozygous and 50 to be heterozygous. The 50 homozygous loci of interest, or the 50 heterozygous loci of interest or the 50 homozygous and 50 heterozygous loci of interest, or any combination of the homozygous and heterozygous loci of interest on each chromosome can be analyzed using the template DNA from the sample from the pregnant female.

The locus of interest on the template DNA from the sample of the pregnant female is analyzed using the amplification, isolation, digestion, fill in, and detection methods described above. The same primers used to analyze the locus of interest on the maternal template DNA are used to screen the template DNA from the sample from the pregnant female. Any number of loci of interest can be analyzed on the template DNA from the sample from the pregnant female. For example, 1, 1-5, 5-10, 10-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90, 90-100, 100-150, 150-200, 200-250, 250-300, 300-500, 500-1000, 1000-2000, 2000-3000, 3000-4000 or more than 4000 homozygous maternal loci of interest can be analyzed in the template DNA from the sample from the pregnant female. In a preferred embodiment, multiple loci of interest on multiple chromosomes are analyzed.

From the population of homozygous maternal loci of interest, there will be both heterozygous and homozygous loci of interest from the template DNA from the sample from the pregnant female; the heterozygous loci of interest can be further analyzed. At heterozygous loci of interest, the ratio of alleles can be used to determine the number of chromosomes that are present.

The percentage of fetal DNA present in the sample from the pregnant female can be calculated by determining the ratio of alleles at a heterozygous locus of interest on a chromosome that is not typically associated with a chromosomal abnormality. In a preferred embodiment, the ratio of alleles at multiple heterozygous loci of interest on a chromosome can be used to determine the percentage of fetal DNA. For example, chromosome 1, which is the largest chromosome in the human genome, can be used to determine the percentage of fetal DNA.

For example, suppose SNP X is homozygous at the maternal template DNA (A/A). At SNP X, the template DNA from the sample from the pregnant female, which can contain both fetal DNA and maternal DNA, is heterozygous (A/G). The nucleotide guanine represents the fetal DNA because at SNP X the mother is homozygous, and thus the guanine is attributed to the fetal DNA. The guanine at SNP X can be used to calculate the percentage of fetal DNA in the sample.

Alternatively, multiple loci of interest on two or more chromosomes can be examined to determine the percentage of fetal DNA. For example, multiple loci of interest can be examined on chromosomes 13, and 18 to determine the percentage of fetal DNA because organisms with chromosomal abnormalities at chromosome 13 and 18 are not viable.

Alternatively, for a male fetus, a marker on the Y chromosome can be used to determine the amount of fetal DNA present in the sample. A panel of serial dilutions can be made using the template DNA isolated from the sample from the pregnant female, and quantitative PCR analysis performed. Two PCR reactions can be performed: one PCR reaction to amplify a marker on the Y chromosome, for example SRY, and the other reaction to amplify a region on any of the autosomal chromosomes. The amount of fetal DNA can be calculated using the following formula:

Percent Fetal DNA: (last dilution Y chromosome detected / last dilution autosomal chromosome detected) *2 * 100.

The expected ratio of the paternal allele to the maternal allele depends on the amount of fetal DNA present in the sample from the pregnant female. For example, if at SNP A, the mother is homozygous A/A, and the fetus is heterozygous A/G, then the ratio of A:G can be used to detect chromosomal abnormalities. If the fetal DNA is fifty percent (50%) of the DNA in the maternal blood, then at SNP A where the maternal nucleotide is an adenine and the other nucleotide, which is contributed by the father, is a guanine, one would expect the ratio of adenine (two adenines from the maternal template DNA and one from the fetal template DNA) to guanine (from the fetal template DNA) to be 75:25. However, if the fetus has a trisomy of this particular chromosome, and the additional chromosome is contributed by the mother, and thus an additional adenine nucleotide is present, then one would expect the ratio of 83.4:16.6 (the fetal DNA is 50% of the DNA in the maternal blood, so each nucleotide contributed by the fetus, the two adenines and the guanine, are each 16.66% of the total DNA in the sample). Thus, an 8%

increase in the signal for adenine and an 8% decrease in the signal for guanine would be detected. On the other hand, if the additional chromosome is contributed by the father, and thus, an additional guanine is present, then one would expect the ratio of 66.6:33.4.

However, if the fetal DNA is 40% of the DNA in the maternal blood, the expected ratio without a trisomy is 80:20. If the fetus has a trisomy, and the additional chromosome is provided by the mother, the expected ratio would be 86.6:13.3. A 6.6% increase in signal for the adenine and a 6.6 % decrease in the signal for guanine would be detected.

In another embodiment, multiple loci of interest on multiple chromosomes can be examined. The ratios for the alleles at each heterozygous locus of interest on a chromosome can be summed and compared to the ratios for the alleles at each locus of interest on a different chromosome. The chromosomes that are compared can be of human origin, and include but are not limited to chromosomes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, and Y. The ratio obtained from multiple chromosomes can be compared to the ratio obtained for a single chromosome or from multiple chromosomes.

In one embodiment, one of the chromosomes used in the comparison can be chromosome 13, 16; 18, 21, 22, X or Y. In a preferred embodiment, the ratios on chromosomes 13, 18, and 21 are compared.

For example, assuming 40% fetal DNA in the sample from the pregnant female, the ratio of the alleles at a heterozygous locus of interest on chromosome 1 will be 80:20. Likewise, the ratio of alleles at a heterozygous locus of interest on chromosome 21 will be present in a ratio of 80:20. However, in a fetus with trisomy 21 where the additional chromosome is contributed by the mother, the nucleotides at a heterozygous locus of interest on chromosome 21 will be present in a ratio of 86.6:13.3. By contrast, the ratio for chromosome 1 will remain at 80:20, and thus the 6.6% increase in the maternal nucleotide at chromosome 21 will signify an additional chromosome or part of a chromosome. One to tens to hundreds to thousands of loci of interest can be analyzed.

In another embodiment, the loci of interest on the template DNA from the sample from the pregnant female can be genotyped; heterozygous and homozygous loci of interest will be identified. The ratio of the alleles at the loci of interest can be used to determine the presence or absence of a chromosomal abnormality. The template DNA from the sample from the pregnant female contains both maternal template DNA and

fetal template DNA. There are 3 possibilities at each SNP for either the maternal template DNA or the fetal template DNA: heterozygous, homozygous for allele 1, or homozygous for allele 2. The possible nucleotide ratios for a SNP that is either an adenine or a guanine are shown in Table II. The ratios presented in Table II are calculated with the fetal DNA at 50% of the DNA in the sample from the pregnant female.

Table 11. Ratios for nucleotides for a heterozygous SNP.

	Fetal SNP		
Maternal SNP	A/A	G/G	A/G
A/A	100% A	N/A	75% A, 25%G
G/G	N/A	100% G	25% A,75% G
A/G	75% A, 25%G	25% A, 75% G	50% A, 50% G

There are three nucleotide ratios: 100% of a single nucleotide, 50:50, or 75:25. These ratios will vary depending on the amount of fetal DNA present in sample from the pregnant female. However, the percentage of fetal DNA should be constant regardless of the chromosome analyzed. Therefore, if chromosomes are present in two copies, the above calculated ratios will be seen.

On the other hand, these percentages will vary when an additional chromosome is present. For example, assume that SNP X can be adenine or guanine, and that the percentage of fetal DNA in the sample from the pregnant female is 50%. Analysis of the loci of interest on chromosome 1 will provide the ratios discussed above: 100:0, 50:50, and 75:25. The possible ratios for a SNP that is A/G with an additional chromosome are provided in Table III.

Table III: Nucleotides ratios at a SNP when an additional copy of a chromosome is present

	Fetal SNP			
Maternal SNPX	A/A/A	G/G/G	A/G/G	A/A/G
A/A	100% A	N/A	60% A, 40%G	80% A, 20%G
G/G	N/A	100% G	20% A, 80% G	40% A, 60% G
A/G	80% A, 20% G	20% A, 80% G	40% A, 60% G	60% A, 40% G

The possible ratios for the alleles at a heterozygous SNP with an additional copy of a chromosome are: 100:0, 60:40, and 80:20. Two of these ratios, 60:40, and 80:20 differ from the ratios of alleles at heterozygous SNPs obtained with two copies of a chromosome. As discussed above, the ratios for the nucleotides at a heterozygous SNP depend on the amount of fetal DNA present in the sample. However, the ratios, whatever they are, will remain constant across chromosomes unless there is a chromosomal abnormality.

The ratio of alleles at heterozygous loci of interest on a chromosome can be compared to the ratio for alleles at heterozygous loci of interest on a different chromosome. For example, the ratio for multiple loci of interest on chromosome 1 (the ratio at SNP 1, SNP 2, SNP 3, SNP 4, etc.) can be compared to the ratio for multiple loci of interest on chromosome 21 (the ratio at SNP A, SNP B, SNP C, SNP D, etc.). Any chromosome can be compared to any other chromosome. There is no limit to the number of chromosomes that can be compared.

Referring back to the data in Tables II and III, the ratios for nucleotides at a heterozygous SNP on chromosome 1, which was present in two copies, were 75:25, and 50:50. On the other, the ratio for nucleotides at a heterozygous SNP on chromosome 21, which was present in three copies, were 60:40, and 80:20. The difference between these two ratios indicates a chromosomal abnormality. The ratios can be pre-calculated for the full range of varying degrees of fetal DNA present in the maternal serum. Tables II and III demonstrate that both maternal homozygous and heterozygous loci of interest can be used to detect the presence of a fetal chromosomal abnormality.

The above example illustrates how the ratios for nucleotides at heterozygous SNPs can be used to detect the presence of an additional chromosome. The same type of analysis can be used to detect chromosomal rearrangements, translocations, mini-chromosomes, duplications of regions of chromosomes, monosomies, deletions of regions of chromosomes, and fragments of chromosomes. The present invention does not quantitate the amount of a fetal gene product, nor is the utility of the present invention limited to the analysis of genes found on the Y chromosome. The present invention does not merely rely on the detection of a paternally inherited nucleic acid, rather, the present invention provides a method that allows the ratio of maternally to paternally inherited

alleles at loci of interest, including SNPs, to be calculated. The method does not require genotyping of the mother or the father.

Any chromosome of any organism can be analyzed using the methods of the invention. For example, in humans, chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X or Y can be analyzed using the methods of the invention. The ratio for the alleles at a heterozygous locus of interest on any chromosome can be compared to the ratio for the alleles at a heterozygous locus of interest on any other chromosome.

Thus, the present invention provides a non-invasive technique, which is independent of fetal cell isolation, for rapid, accurate and definitive detection of chromosome abnormalities in a fetus. The present invention also provides a non-invasive method for determining the sequence of DNA from a fetus. The present invention can be used to detect any alternation in gene sequence as compared to the wild type sequence including but not limited to point mutation, reading frame shift, transition, transversion, addition, insertion, deletion, addition-deletion, frame-shift, missense, reverse mutation, and microsatellite alteration.

Detection of Fetal Chromosomal Abnormalities Using Short Tandem Repeats

Short tandem repeats (STRs) are short sequences of DNA, normally of 2-5 base pairs in length, which are repeated numerous times in a head-tail manner. Tandemly repeated DNA sequences are widespread throughout the human genome, and show sufficient variability among the individuals in a population. Minisatellites have core repeats with 9-80 base pairs.

In another embodiment, short tandem repeats can be used to detect fetal chromosomal abnormalities. Template DNA can be obtained from a nucleic acid containing sample including but not limited to cell, tissue, blood, serum, plasma, saliva, urine, tears, vaginal secretion, lymph fluid, cerebrospinal fluid, mucosa secretion, peritoneal fluid, ascitic fluid, fecal matter, or body exudates. In another embodiment, a cell lysis inhibitor is added to the nucleic acid containing sample. In a preferred embodiment, the template DNA is obtained from the blood of a pregnant female. In another embodiment, the template DNA is obtained from the plasma or serum from the blood of a pregnant female.

The template DNA obtained from the blood of the pregnant female will contain both fetal DNA and maternal DNA. The fetal DNA comprises STRs from the mother and

the father. The variation in the STRs between the mother and father can be used to detect chromosomal abnormalities.

Primers can be designed to amplify short tandem repeats. Any method of amplification can be used including but not limited to polymerase chain reaction, self-sustained sequence reaction, ligase chain reaction, rapid amplification of cDNA ends, polymerase chain reaction and ligase chain reaction, Q-beta phage amplification, strand displacement amplification, and splice overlap extension polymerase chain reaction. In a preferred embodiment, PCR is used.

Any number of short tandem repeats can be analyzed including but not limited to 1-5, 5-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-1000, and greater than 1000. The short tandem repeats can be analyzed in a single PCR reaction or in multiple PCR reactions. In a preferred embodiment, STRs from multiple chromosomes are analyzed.

After amplification, the PCR products can be analyzed by any number of methods including but not restricted to gel electrophoresis, and mass spectrometry. The template DNA from the pregnant female comprises STRs of maternal and paternal origin. The STRs of paternal origin represent the fetal DNA. The paternal and maternal STRs may be identical in length or the maternal and the paternal STRs may differ.

Heterozygous STRs are those of which the maternal and paternal differ in length. The amount of each PCR product can be quantitated for each heterozygous STR. With a normal number of chromosomes, the amount of each PCR product should be approximately equal. However, with an extra chromosome, one of the STR PCR products will be present at a greater amount.

For example, multiple STRs on chromosome 1 can be analyzed on the template DNA obtained from the blood of the pregnant female. Each STR, whether of maternal or paternal origin, should be present at approximately the same amount. Likewise, with two chromosome 21s, each STR should be present at approximately the same amount. However, with a trisomy 21, one of the STR PCR products, when the maternal and paternal differ in length (a heterozygous STR) should be present at a higher amount. The ratio for each heterozygous STR on one chromosome can be compared to the ratio for each heterozygous STR on a different chromosome, wherein a difference indicates the presence or absence of a chromosomal abnormality.

Kits

The methods of the invention are most conveniently practiced by providing the reagents used in the methods in the form of kits. A kit preferably contains one or more of the following components: written instructions for the use of the kit, appropriate buffers, salts, DNA extraction detergents, primers, nucleotides, labeled nucleotides, 5' end modification materials, and if desired, water of the appropriate purity, confined in separate containers or packages, such components allowing the user of the kit to extract the appropriate nucleic acid sample, and analyze the same according to the methods of the invention. The primers that are provided with the kit will vary, depending upon the purpose of the kit and the DNA that is desired to be tested using the kit.

A kit can also be designed to detect a desired or variety of single nucleotide polymorphisms, especially those associated with an undesired condition or disease. For example, one kit can comprise, among other components, a set or sets of primers to amplify one or more loci of interest associated with Huntington's disease. Another kit can comprise, among other components, a set or sets of primers for genes associated with a predisposition to develop type I or type II diabetes. Still, another kit can comprise, among other components, a set or sets of primers for genes associated with a predisposition to develop heart disease. Details of utilities for such kits are provided in the "Utilities" section below.

Utilities

The methods of the invention can be used whenever it is desired to know the genotype of an individual. The method of the invention is especially useful for the detection of genetic disorders. The method of the invention is especially useful as a non-invasive technique for the detection of genetic disorders in a fetus. In a preferred embodiment, the method of the invention provides a method for identification of single nucleotide polymorphisms.

In a preferred embodiment, the method is useful for detecting chromosomal abnormalities including but not limited to trisomies, monosomies, duplications, deletions, additions, chromosomal rearrangements, translocations, and other anueplodies. The method is especially useful for the detection of chromosomal abnormalities in a fetus.

In a preferred embodiment, the method of the invention provides a method for identification of the presence of a disease in a fetus, especially a genetic disease that arises as a result of the presence of a genomic sequence, or other biological condition that

it is desired to identify in an individual for which it is desired to know the same. The identification of such sequence in the fetus based on the presence of such genomic sequence can be used, for example, to determine if the fetus is a carrier or to assess if the fetus is predisposed to developing a certain genetic trait, condition or disease. The method of the invention is especially useful in prenatal genetic testing of parents and child.

Examples of diseases that can be diagnosed by this invention are listed in Table IV.

TABLE IV

Achondroplasia
Adrenoleukodystrophy, X-Linked
Agammaglobulinemia, X-Linked
Alagille Syndrome
Alpha-Thalassemia X-Linked Mental Retardation Syndrome
Alzheimer Disease
Alzheimer Disease, Early-Onset Familial
Amyotrophic Lateral Sclerosis Overview
Androgen Insensitivity Syndrome
Angelman Syndrome
Ataxia Overview, Hereditary
Ataxia-Telangiectasia
Becker Muscular Dystrophy also The Dystrophinopathies)
Beckwith-Wiedemann Syndrome
Beta-Thalassemia
Biotinidase Deficiency
Branchiootorenal Syndrome
BRCA1 and BRCA2 Hereditary Breast/Ovarian Cancer
Breast Cancer
CADASIL
Canavan Disease
Cancer

Charcot-Marie-Tooth Hereditary Neuropathy
Charcot-Marie-Tooth Neuropathy Type 1
Charcot-Marie-Tooth Neuropathy Type 2
Charcot-Marie-Tooth Neuropathy Type 4
Charcot-Marie-Tooth Neuropathy Type X
Cockayne Syndrome
Colon Cancer
Contractural Arachnodactyly, Congenital
Craniosynostosis Syndromes (FGFR-Related)
Cystic Fibrosis
Cystinosis
Deafness and Hereditary Hearing Loss
DRPLA (Dentatorubral-Pallidoluysian Atrophy)
DiGeorge Syndrome (also 22q11 Deletion Syndrome)
Dilated Cardiomyopathy, X-Linked
Down Syndrome (Trisomy 21)
Duchenne Muscular Dystrophy (also The Dystrophinopathies)
Dystonia, Early-Onset Primary (DYT1)
Dystrophinopathies, The
Ehlers-Danlos Syndrome, Kyphoscoliotic Form
Ehlers-Danlos Syndrome, Vascular Type
Epidermolysis Bullosa Simplex
Exostoses, Hereditary Multiple
Facioscapulohumeral Muscular Dystrophy
Factor V Leiden Thrombophilia
Familial Adenomatous Polyposis (FAP)
Familial Mediterranean Fever
Fragile X Syndrome
Friedreich Ataxia
Frontotemporal Dementia with Parkinsonism-17
Galactosemia

Gaucher Disease
Hemochromatosis, Hereditary
Hemophilia A
Hemophilia B
Hemorrhagic Telangiectasia, Hereditary
Hearing Loss and Deafness, Nonsyndromic, DFNA (Connexin 26)
Hearing Loss and Deafness, Nonsyndromic, DFNB 1 (Connexin 26)
Hereditary Spastic Paraplegia
Hermansky-Pudlak Syndrome
Hexosaminidase A Deficiency (also Tay-Sachs)
Huntington Disease
Hypochondroplasia
Ichthyosis, Congenital, Autosomal Recessive
Incontinentia Pigmenti
Kennedy Disease (also Spinal and Bulbar Muscular Atrophy)
Krabbe Disease
Leber Hereditary Optic Neuropathy
Lesch-Nyhan Syndrome Leukemias
Li-Fraumeni Syndrome
Limb-Girdle Muscular Dystrophy
Lipoprotein Lipase Deficiency, Familial
Lissencephaly
Marfan Syndrome
MELAS (Mitochondrial Encephalomyopathy, Lactic Acidosis, and Stroke-Like Episodes)
Monosomies
Multiple Endocrine Neoplasia Type 2
Multiple Exostoses, Hereditary Muscular Dystrophy, Congenital
Myotonic Dystrophy
Nephrogenic Diabetes Insipidus
Neurofibromatosis 1

Neurofibromatosis 2
Neuropathy with Liability to Pressure Palsies, Hereditary
Niemann-Pick Disease Type C
Nijmegen Breakage Syndrome Norrie Disease
Oculocutaneous Albinism Type 1
Oculopharyngeal Muscular Dystrophy
Ovarian Cancer
Pallister-Hall Syndrome
Parkin Type of Juvenile Parkinson Disease
Pelizaeus-Merzbacher Disease
Pendred Syndrome
Peutz-Jeghers Syndrome Phenylalanine Hydroxylase Deficiency
Prader-Willi Syndrome
PROP 1 -Related Combined Pituitary Hormone Deficiency (CPHD)
Prostate Cancer
Retinitis Pigmentosa
Retinoblastoma
Rothmund-Thomson Syndrome
Smith-Lemli-Opitz Syndrome
Spastic Paraplegia, Hereditary
Spinal and Bulbar Muscular Atrophy (also Kennedy Disease)
Spinal Muscular Atrophy
Spinocerebellar Ataxia Type 1
Spinocerebellar Ataxia Type 2
Spinocerebellar Ataxia Type 3
Spinocerebellar Ataxia Type 6
Spinocerebellar Ataxia Type 7
Stickler Syndrome (Hereditary Arthroophthalmopathy)
Tay-Sachs (also GM2 Gangliosidoses)
Trisomies
Tuberous Sclerosis Complex

Usher Syndrome Type I
Usher Syndrome Type II
Velocardiofacial Syndrome (also 22q11 Deletion Syndrome)
Von Hippel-Lindau Syndrome
Williams Syndrome
Wilson Disease
X-Linked Adrenoleukodystrophy
X-Linked Agammaglobulinemia
X-Linked Dilated Cardiomyopathy (also The Dystrophinopathies)
X-Linked Hypotonic Facies Mental Retardation Syndrome

The method of the invention is useful for screening an individual at multiple loci of interest, such as tens, hundreds, or even thousands of loci of interest associated with a genetic trait or genetic disease by sequencing the loci of interest that are associated with the trait or disease state, especially those most frequently associated with such trait or condition. The invention is useful for analyzing a particular set of diseases including but not limited to heart disease, cancer, endocrine disorders, immune disorders, neurological disorders, musculoskeletal disorders, ophthalmologic disorders, genetic abnormalities, trisomies, monosomies, transversions, translocations, skin disorders, and familial diseases.

The method of the invention can also be used to confirm or identify the relationship of a DNA of unknown sequence to a DNA of known origin or sequence, for example, for use in, maternity or paternity testing, and the like.

Having now generally described the invention, the same will become better understood by reference to certain specific examples which are included herein for purposes of illustration only and are not intended to be limiting unless other wise specified.

EXAMPLES

The following examples are illustrative only and are not intended to limit the scope of the invention as defined by the claims.

EXAMPLE 1

DNA sequences were amplified by PCR, wherein the annealing step in cycle 1 was performed at a specified temperature, and then increased in cycle 2, and further increased in cycle 3 for the purpose of reducing non-specific amplification. The TM1 of cycle 1 of PCR was determined by calculating the melting temperature of the 3' region, which anneals to the template DNA, of the second primer. For example, in FIG. 1B, the TM1 can be about the melting temperature of region "c." The annealing temperature was raised in cycle 2, to TM2, which was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. For example, in FIG. 1C, the annealing temperature (TM2) corresponds to the melting temperature of region "b." In cycle 3, the annealing temperature was raised to TM3, which was about the melting temperature of the entire sequence of the second primer For example, in FIG. 1D, the annealing temperature (TM3) corresponds to the melting temperature of region "c" + region "d". The remaining cycles of amplification were performed at TM3.

Preparation of Template DNA

The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Blood was collected from 36 volunteers. Template DNA was isolated from each blood sample using QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). Following isolation, the template DNA from each of the 36 volunteers was pooled for further analysis.

Primer Design

The following four single nucleotide polymorphisms were analyzed: SNP HC21S00340, identification number as assigned by Human Chromosome 21 cSNP Database, (FIG. 3, lane 1) located on chromosome 21; SNP TSC 0095512 (FIG. 3, lane 2) located on chromosome 1, SNP TSC 0214366 (FIG. 3, lane 3) located on chromosome 1; and SNP TSC 0087315 (FIG. 3, lane 4) located on chromosome 1. The SNP Consortium Ltd database can be accessed at http://snp.cshl.org/, website address effective as of February 14, 2002.

SNP HC21S00340 was amplified using the following primers:

First primer:

5'TAGAATAGCACTGAATTCAGGAATACAATCATTGTCAC 3'

Second primer:

5'ATCACGATAAACGGCCAAACTCAGGTTA3'

SNP TSC0095512 was amplified using the following primers:

First primer:

5'AAGTTTAGATCAGAATTCGTGAAAGCAGAAGTTGTCTG 3'

Second primer:

5'TCTCCAACTAACGGCTCATCGAGTAAAG 3'

SNP TSC0214366 was amplified using the following primers:

First primer:

5'ATGACTAGCTATGAATTCGTTCAAGGTAGAAAATGGAA 3'

Second primer:

5'GAGAATTAGAACGGCCCAAATCCCACTC3'

SNP TSC 0087315 was amplified using the following primers:

First primer:

5'TTACAATGCATGAATTCATCTTGGTCTCTCAAAGTGC 3'

Second primer:

5'TGGACCATAAACGGCCAAAAACTGTAAG 3'.

All primers were designed such that the 3' region was complementary to either the upstream or downstream sequence flanking each locus of interest and the 5' region contained a restriction enzyme recognition site. The first primer contained a biotin tag at the 5' end and a recognition site for the restriction enzyme EcoRI. The second primer contained the recognition site for the restriction enzyme BceA I.

PCR Reaction

All four loci of interest were amplified from the template genomic DNA using PCR (U.S. Patent Nos. 4,683,195 and 4,683,202). The components of the PCR reaction were as follows: 40 ng of template DNA, 5 μM first primer, 5 μM second primer, 1 X HotStarTaq Master Mix as obtained from Qiagen (Catalog No. 203443). The HotStarTaq Master Mix contained DNA polymerase, PCR buffer, 200 μM of each dNTP, and 1.5 mM $MgCl_2$.

Amplification of each template DNA that contained the SNP of interest was performed using three different series of annealing temperatures, herein referred to as low stringency annealing temperature, medium stringency annealing temperature, and high stringency annealing temperature. Regardless of the annealing temperature protocol, each PCR reaction consisted of 40 cycles of amplification. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN. As instructed by

the manufacturer, the reactions were incubated at 95°C for 15 min. prior to the first cycle of PCR. The denaturation step after each extension step was performed at 95°C for 30 sec. The annealing reaction was performed at a temperature that permitted efficient extension without any increase in temperature.

The low stringency annealing reaction comprised three different annealing temperatures in each of the first three cycles. The annealing temperature for the first cycle was 37°C for 30 sec.; the annealing temperature for the second cycle was 57°C for 30 sec.; the annealing temperature for the third cycle was 64°C for 30 sec. Annealing was performed at 64°C for subsequent cycles until completion.

As shown in the photograph of the gel (FIG. 3A), multiple bands were observed after amplification of SNP TSC 0087315 (lane 4). Amplification of SNP HC21S00340 (lane 1), SNP TSC0095512 (lane 2), and SNP TSC0214366 (lane 3) generated a single band of high intensity and one band of faint intensity, which was of higher molecular weight. When the low annealing temperature conditions were used, the correct size product was generated and this was the predominant product in each reaction.

The medium stringency annealing reaction comprised three different annealing temperatures in each of the first three cycles. The annealing temperature for the first cycle was 40°C for 30 seconds; the annealing temperature for the second cycle was 60°C for 30 seconds; and the annealing temperature for the third cycle was 67°C for 30 seconds. Annealing was performed at 67°C for subsequent cycles until completion. Similar to what was observed under low stringency annealing conditions, amplification of SNP TSC0087315 (FIG. 3B, lane 4) generated multiple bands under conditions of medium stringency. Amplification of the other three SNPs (lanes 1-3) produced a single band. These results demonstrate that variable annealing temperatures can be used to cleanly amplify loci of interest from genomic DNA with a primer that has an annealing length of 13 bases.

The high stringency annealing reaction was comprised of three different annealing temperatures in each of the first three cycles. The annealing temperature of the first cycle was 46°C for 30 seconds; the annealing temperature of the second cycle was 65°C for 30 seconds; and the annealing temperature for the third cycle was 72°C for 30 seconds. Annealing was performed at 72°C for subsequent cycles until completion. As shown in the photograph of the gel (FIG. 3C), amplification of SNP TSC0087315 (lane 4) using the high stringency annealing temperatures generated a single band of the correct

molecular weight. By raising the annealing temperatures for each of the first three cycles, non-specific amplification was eliminated. Amplification of SNP TSC0095512 (lane 2) generated a single band. SNPs HC21S00340 (lane 1), and TSC0214366 (lane 3) failed to amplify at the high stringency annealing temperatures, however, at the medium stringency annealing temperatures, these SNPs amplified as a single band. These results demonstrate that variable annealing temperatures can be used to reduce non-specific PCR products, as demonstrated for SNP TSC0087315 (FIG. 3, lane 4).

EXAMPLE 2

SNPs on chromosomes 1 (TSC0095512), 13 (TSC0264580), and 21 (HC21S00027) were analyzed. SNP TSC0095512 was analyzed using two different sets of primers, and SNP HC21S00027 was analyzed using two types of reactions for the incorporation of nucleotides.

Preparation of Template DNA

The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Template DNA was isolated using the QIAmp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). The template DNA was isolated as per instructions included in the kit. Following isolation, template DNA from thirty-six human volunteers were pooled together and cut with the restriction enzyme EcoRI. The restriction enzyme digestion was performed as per manufacturer's instructions.

Primer Design

SNP HC21S00027 was amplified by PCR using the following primer set:

First primer:

5' ATAACCGTATGCGAATTCTATAATTTTCCTGATAAAGG 3'

Second primer:

5' CTTAAATCAGGGGACTAGGTAAACTTCA 3'.

The first primer contained a biotin tag at the extreme 5' end, and the nucleotide sequence for the restriction enzyme EcoRI. The second primer contained the nucleotide sequence for the restriction enzyme BsmF I (FIG. 4A).

Also, SNP HC21S00027 was amplified by PCR using the same first primer but a different second primer with the following sequence:

Second primer:

5' CTTAAATCAGACGGCTAGGTAAACTTCA 3'

This second primer contained the recognition site for the restriction enzyme BceA I (FIG. 4B).

SNP TSC0095512 was amplified by PCR using the following primers:

First primer:

5' AAGTTTAGATCAGAATTCGTGAAAGCAGAAGTTGTCTG 3'

Second primer:

5' TCTCCAACTAGGGACTCATCGAGTAAAG 3'.

The first primer had a biotin tag at the 5' end and contained a restriction enzyme recognition site for EcoRI. The second primer contained a restriction enzyme recognition site for BsmF I (FIG. 4C).

Also, SNP TSCO095512 was amplified using the same first primer and a different second primer with the following sequence:

Second primer:

5'TCTCCAACTAACGGCTCATCGAGTAAAG3'

This second primer contained the recognition site for the restriction enzyme BceA I (FIG. 4D).

SNP TSC0264580, which is located on chromosome 13, was amplified with the following primers:

First primer:

5' AACGCCGGGCGAGAATTCAGTTTTTCAACTTGCAAGG 3'

Second primer:

5' CTACACATATCTGGGACGTTGGCCATCC 3'.

The first primer contained a biotin tag at the extreme 5' end and had a restriction enzyme recognition site for EcoRI. The second primer contained a restriction enzyme recognition site for BsmF I.

PCR Reaction

All loci of interest were amplified from the template genomic DNA using the polymerase chain reaction (PCR, U.S. Patent Nos. 4,683,195 and 4,683,202, incorporated herein by reference). In this example, the loci of interest were amplified in separate reaction tubes but they could also be amplified together in a single PCR reaction. For increased specificity, a "hot-start" PCR was used. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN (catalog number 203443). The

amount of template DNA and primer per reaction can be optimized for each locus of interest but in this example, 40 ng of template human genomic DNA and 5 μM of each primer were used. Forty cycles of PCR were performed. The following PCR conditions were used:

(1) 95°C for 15 minutes and 15 seconds;
(2) 37°C for 30 seconds;
(3) 95°C for 30 seconds;
(4) 57°C for 30 seconds;
(5) 95°C for 30 seconds;
(6) 64°C for 30 seconds;
(7) 95°C for 30 seconds;
(8) Repeat steps 6 and 7 thirty nine (39) times;
(9) 72°C for 5 minutes.

In the first cycle of PCR, the annealing temperature was about the melting temperature of the 3' annealing region of the second primers, which was 37°C. The annealing temperature in the second cycle of PCR was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer, which was 57°C. The annealing temperature in the third cycle of PCR was about the melting temperature of the entire sequence of the second primer, which was 64°C. The annealing temperature for the remaining cycles was 64°C. Escalating the annealing temperature from TM1 to TM2 to TM3 in the first three cycles of PCR greatly improves specificity. These annealing temperatures are representative, and the skilled artisan will understand the annealing temperatures for each cycle are dependent on the specific primers used.

The temperatures and times for denaturing, annealing, and extension, can be optimized by trying various settings and using the parameters that yield the best results. The PCR products for SNP HC21S00027 and SNP TSC095512 are shown in FIGS. 5A-5D.

Purification of Fragment of Interest

The PCR products were separated from the genomic template DNA. Each PCR product was divided into four separate reaction wells of a Streptawell, transparent, High-Bind plate from Roche Diagnostics GmbH (catalog number 1 645 692, as listed in Roche Molecular Biochemicals, 2001 Biochemicals Catalog). The first primers contained a 5' biotin tag so the PCR products bound to the Streptavidin coated wells

while the genomic template DNA did not. The streptavidin binding reaction was performed using a Thermomixer (Eppendorf) at 1000 rpm for 20 min. at 37°C. Each well was aspirated to remove unbound material, and washed three times with 1X PBS, with gentle mixing (Kandpal et al., Nucl. Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucl. Acids Res. 18:6163-6164 (1990)).

Restriction Enzyme Digestion of Isolated Fragments

The purified PCR products were digested with the restriction enzyme that bound the recognition site incorporated into the PCR products from the second primer. SNP HC21S00027 (FIG. 6A and 6B) and SNP TSC0095512 (FIG. 6C and 6D) were amplified in separate reactions using two different second primers. FIG. 6A (SNP HC21S00027) and FIG. 6C (SNP TSC0095512) depict the PCR products after digestion with the restriction enzyme BsmF I (New England Biolabs catalog number R0572S). FIG. 6B (SNP HC21S00027) and FIG. 6D (SNP TSC0095512) depict the PCR products after digestion with the restriction enzyme BceA I (New England Biolabs, catalog number R0623 S). The digests were performed in the Streptawells following the instructions supplied with the restriction enzyme. SNP TSC0264580 was digested with BsmF I. After digestion with the appropriate restriction enzyme, the wells were washed three times with PBS to remove the cleaved fragments.

Incorporation of Labeled Nucleotide

The restriction enzyme digest described above yielded a DNA fragment with a 5' overhang, which contained the SNP site or locus of interest and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide or nucleotides in the presence of a DNA polymerase.

For each SNP, four separate fill in reactions were performed; each of the four reactions contained a different fluorescently labeled ddATP (ddATP, ddATP, ddATP, or ddATP). The following components were added to each fill in reaction: 1 μl of a fluorescently labeled ddATP, 0.5 μl of unlabeled ddNTPs (40 μM), which contained all nucleotides except the nucleotide that was fluorescently labeled, 2 μl of 10X sequenase buffer, 0.25 μl of Sequenase, and water as needed for a 20 μl reaction. All of the fill in reactions were performed at 40°C for 10 min. Non-fluorescently labeled ddATP was purchased from Fermentas Inc. (Hanover, MD). All other labeling reagents were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565). In the presence of fluorescently labeled ddNTPs, the 3' recessed end

was extended by one base, which corresponds to the SNP or locus of interest (FIG 7A-7D).

A mixture of labeled ddNTPs and unlabeled dNTPs also was used for the "fill in" reaction for SNP HC21S00027. The "fill in" conditions were as described above except that a mixture containing 40 μM unlabeled dNTPs, 1 μl fluorescently labeled ddATP, 1 μl fluorescently labeled ddATP, 1 μl fluorescently labeled ddATP, and 1 μl ddATP was used. The fluorescent ddNTPs were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit, US 79565; Amersham did not publish the concentrations of the fluorescent nucleotides). SNP HC21S00027 was digested with the restriction enzyme BsmF I, which generated a 5' overhang of four bases. As shown in FIG. 7E, if the first nucleotide incorporated is a labeled ddATP, the 3' recessed end is filled in by one base, allowing detection of the SNP or locus of interest. However, if the first nucleotide incorporated is a dNTP, the polymerase continues to incorporate nucleotides until a ddNTP is filled in. For example, the first two nucleotides can be filled in with dNTPs, and the third nucleotide with a ddNTP, allowing detection of the third nucleotide in the overhang. Thus, the sequence of the entire 5' overhang can be determined, which increases the information obtained from each SNP or locus of interest.

After labeling, each Streptawell was rinsed with 1X PBS (100 μl) three times. The "filled in" DNA fragments were then released from the Streptawells by digestion with the restriction enzyme EcoRI, according to the manufacturer's instructions that were supplied with the enzyme (FIGS. 8A-8D). Digestion was performed for 1 hour at 37°C with shaking at 120 rpm.

Detection of the Locus of Interest

After release from the streptavidin matrix, 2-3 μl of the 10 μl sample was loaded in a 48 well membrane tray (The Gel Company, catalog number TAM48-01). The sample in the tray was absorbed with a 48 Flow Membrane Comb (The Gel Company, catalog number AM48), and inserted into a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691).

The sample was electrophoresed into the gel at 3000 volts for 3 min. The membrane comb was removed, and the gel was run for 3 hours on an ABI 377 Automated Sequencing Machine. The incorporated labeled nucleotide was detected by fluorescence.

As shown in FIG. 9A, from a sample of thirty six (36) individuals, one of two nucleotides, either adenosine or guanine, was detected at SNP HC21S00027. These are the two nucleotides reported to exist at SNP HC21S00027 (http://snp.cshl.org/snpsearch.shtml).

One of two nucleotides, either guanine or cytosine, was detected at SNP TS00095512 (FIG. 9B). The same results were obtained whether the locus of interest was amplified with a second primer that contained a recognition site for BceA I or the second primer contained a recognition site for BsmF I.

As shown in FIG. 9C, one of two nucleotides was detected at SNP TSC0264580, which was either adenosine or cytosine. These are the two nucleotides reported for this SNP site (http://snp.cshl.org/snpsearch.shtml). In addition, a thymidine was detected one base from the locus of interest. In a sequence dependent manner, BsmF I cuts some DNA molecules at the 10/14 position and other DNA molecules, which have the same sequence, at the 11/15 position. When the restriction enzyme BsmF I cuts 11 nucleotides away on the sense strand and 15 nucleotides away on the antisense strand, the 3' recessed end is one base from the SNP site. The sequence of SNP TSC0264580 indicated that the base immediately preceding the SNP site was a thymidine. The incorporation of a labeled ddNTP into this position generated a fragment one base smaller than the fragment that was cut at the 10/14 position. Thus, the DNA molecules cut at the 11/15 position provided sequence information about the base immediately preceding the SNP site, and the DNA molecules cut at the 10/14 position provided sequence information about the SNP site.

SNP HC21S00027 was amplified using a second primer that contained the recognition site for BsmF I. A mixture of labeled ddNTPs and unlabeled dNTPs was used to fill in the 5' overhang generated by digestion with BsmF I. If a dNTP was incorporated, the polymerase continued to incorporate nucleotides until a ddNTP was incorporated. A population of DNA fragments, each differing by one base, was generated, which allowed the full sequence of the overhang to be determined.

As seen in FIG. 9D, an adenosine was detected, which was complementary to the nucleotide (a thymidine) immediately preceding the SNP or locus of interest. This nucleotide was detected because of the 11/15 cutting property of BsmF I, which is described in detail above. A guanine and an adenosine were detected at the SNP site, which are the two nucleotides reported for this SNP site (FIG. 9A). The two nucleotides

were detected at the SNP site because the molecular weights of the dyes differ, which allowed separation of the two nucleotides. The next nucleotide detected was a thymidine, which is complementary to the nucleotide immediately downstream of the SNP site. The next nucleotide detected was a guanine, which was complementary to the nucleotide two bases downstream of the SNP site. Finally, an adenosine was detected, which was complementary to the third nucleotide downstream of the SNP site. Sequence information was obtained not only for the SNP site but for the nucleotide immediately preceding the SNP site and the next three nucleotides.

None of the loci of interest contained a mutation. However, if one of the loci of interest harbored a mutation including but not limited to a point mutation, insertion, deletion, translocation or any combination of said mutations, it could be identified by comparison to the consensus or published sequence. Comparison of the sequences attributed to each of the loci of interest to the native, non-disease related sequence of the gene at each locus of interest determines the presence or absence of a mutation in that sequence. The finding of a mutation in the sequence is then interpreted as the presence of the indicated disease, or a predisposition to develop the same, as appropriate, in that individual. The relative amounts of the mutated vs. normal or non-mutated sequence can be assessed to determine if the subject has one or two alleles of the mutated sequence, and thus whether the subject is a carrier, or whether the indicated mutation results in a dominant or recessive condition.

EXAMPLE 3

Four loci of interest from chromosome 1 and two loci of interest from chromosome 21 were amplified in separate PCR reactions, pooled together, and analyzed. The primers were designed so that each amplified locus of interest was a different size, which allowed detection of the loci of interest.

Preparation of Template DNA

The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. Template DNA was isolated using the QIAmp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183). The template DNA was isolated as per instructions included in the kit. Template DNA was isolated from thirty-six human volunteers, and then pooled into a single sample for further analysis.

Primer Design

SNP TSC 0087315 was amplified using the following primers:

First primer:

5'TTACAATGCATGAATTCATCTTGGTCTCTCAAAGTGC 3'

Second primer:

5'TGGACCATAAACGGCCAAAAACTGTAAG3'.

SNP TSC0214366 was amplified using the following primers:

First primer:

5'ATGACTAGCTATGAATTCGTTCAAGGTAGAAAATGGAA 3'

Second primer:

5'GAGAATTAGAACGGCCCAAATCCCACTC 3'

SNP TSC 0413944 was amplified with the following primers:

First primer:

5' TACCTTTTGATCGAATTCAAGGCCAAAAATATTAAGTT 3'

Second primer:

5' TCGAACTTTAACGGCCTTAGAGTAGAGA 3'

SNP TSC0095512 was amplified using the following primers:

First primer:

5'AAGTTTAGATCAGAATTCGTGAAAGCAGAAGTTGTCTG 3'

Second primer:

5'TCTCCAACTAACGGCTCATCGAGTAAAG 3'

SNP HC21S00131 was amplified with the following primers:

First primer:

5' CGATTTCGATAAGAATTCAAAAGCAGTTCTTAGTTCAG 3'

Second primer:

5'TGCGAATCTTACGGCTGCATCACATTCA 3'

SNP HC21S00027 was amplified with the following primers:

First primer:

5' ATAACCGTATGCGAATTCTATAATTTTCCTGATAAAGG 3'

Second primer:

5' CTTAAATCAGACGGCTAGGTAAACTTCA 3'

For each SNP, the first primer contained a recognition site for the restriction enzyme EcoRI and had a biotin tag at the extreme 5' end. The second primer used to amplify each SNP contained a recognition site for the restriction enzyme BceA I.

PCR Reaction

The PCR reactions were performed as described in Example 2 except that the following annealing temperatures were used: the annealing temperature for the first cycle of PCR was 37°C for 30 seconds, the annealing temperature for the second cycle of PCR was 57°C for 30 seconds, and the annealing temperature for the third cycle of PCR was 64°C for 30 seconds. All subsequent cycles had an annealing temperature of 64°C for 30 seconds. Thirty seven (37) cycles of PCR were performed. After PCR, 1/4 of the volume was removed from each reaction, and combined into a single tube.

Purification of Fragment of Interest

The PCR products (now combined into one sample, and referred to as "the sample") were separated from the genomic template DNA as described in Example 2 except that the sample was bound to a single well of a Streptawell microtiter plate.

Restriction Enzyme Digestion of Isolated Fragments

The sample was digested with the restriction enzyme BceA I, which bound the recognition site in the second primer. The restriction enzyme digestions were performed following the instructions supplied with the enzyme. After the restriction enzyme digest, the wells were washed three times with 1X PBS.

Incorporation of Nucleotides

The restriction enzyme digest described above yielded DNA molecules with a 5' overhang, which contained the SNP site or locus of interest and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide in the presence of a DNA polymerase.

The following components were used for the fill in reaction: 1 µl of fluorescently labeled ddATP; 1 µl of fluorescently labeled ddTTP; 1 µl of fluorescently labeled ddGTP; 1 µl of fluorescently labeled ddCTP; 2 µl of 10X sequenase buffer, 0.25 µl of Sequenase, and water as needed for a 20 µl reaction. The fill in reaction was performed at 40°C for 10 min. All labeling reagents were obtained from Amersham (Thermo Sequenase Dye Terminator Cycle Sequencing Core Kit (US 79565); the concentration of the ddNTPS provided in the kit is proprietary and not published by Amersham). In the presence of

fluorescently labeled ddNTPs, the 3' recessed end was filled in by one base, which corresponds to the SNP or locus of interest.

After the incorporation of nucleotide, the Streptawell was rinsed with 1X PBS (100 µl) three times. The "filled in" DNA fragments were then released from the Streptawell by digestion with the restriction enzyme EcoRI following the manufacturer's instructions. Digestion was performed for 1 hour at 37 °C with shaking at 120 rpm.

Detection of the Locus of Interest

After release from the streptavidin matrix, 2-3 µl of the 10 µl sample was loaded in a 48 well membrane tray (The Gel Company, catalog number TAM48-01). The sample in the tray was absorbed with a 48 Flow Membrane Comb (The Gel Company, catalog number AM48), and inserted into a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691).

The sample was electrophoresed into the gel at 3000 volts for 3 min. The membrane comb was removed, and the gel was run for 3 hours on an ABI 377 Automated Sequencing Machine. The incorporated nucleotide was detected by fluorescence.

The primers were designed so that each amplified locus of interest differed in size. As shown in FIG. 10, each amplified loci of interest differed by about 5-10 nucleotides, which allowed the loci of interest to be separated from one another by gel electrophoresis. Two nucleotides were detected for SNP TSC0087315, which were guanine and cytosine. These are the two nucleotides reported to exist at SNP TSC0087315 (http://snp.cshl.org/snpsearch.shtml). The sample comprised template DNA from 36 individuals and because the DNA molecules that incorporated a guanine differed in molecular weight from those that incorporated a cytosine, distinct bands were seen for each nucleotide.

Two nucleotides were detected at SNP HC21S00027, which were guanine and adenosine (FIG. 10). The two nucleotides reported for this SNP site are guanine and adenosine (http://snp.cshl.org/snpsearch.shtml). As discussed above, the sample contained template DNA from thirty-six individuals, and one would expect both nucleotides to be represented in the sample. The molecular weight of the DNA fragments that incorporated a guanine was distinct from the DNA fragments that incorporated an adenosine, which allowed both nucleotides to be detected.

The nucleotide cytosine was detected at SNP TSC0214366 (FIG. 10). The two nucleotides reported to exist at this SNP position are thymidine and cytosine.

The nucleotide guanine was detected at SNP TSC0413944 (FIG. 10). The two nucleotides reported for this SNP are guanine and cytosine (http://spp.cshl.org/snpsearch.shtml).

The nucleotide cytosine was detected at SNP TS00095512 (FIG. 10). The two nucleotides reported for this SNP site are guanine and cytosine (http://snp.cshl.org/snpsearch.shtml).

The nucleotide detected at SNP HC21S00131 was guanine. The two nucleotides reported for this SNP site are guanine and adenosine (http://snp.cshl.org/snpsearch.shtml).

As discussed above, the sample was comprised of DNA templates from thirty-six individuals and one would expect both nucleotides at the SNP sites to be represented. For SNP TSC0413944, TSC0095512, TSC0214366 and HC21S00131, one of the two nucleotides was detected. It is likely that both nucleotides reported for these SNP sites are present in the sample but that one fluorescent dye overwhelms the other. The molecular weight of the DNA molecules that incorporated one nucleotide did not allow efficient separation of the DNA molecules that incorporated the other nucleotide. However, the SNPs were readily separated from one another, and for each SNP, a proper nucleotide was incorporated. The sequences of multiple loci of interest from multiple chromosomes, which were treated as a single sample after PCR, were determined.

A single reaction containing fluorescently labeled ddNTPs was performed with the sample that contained multiple loci of interest. Alternatively, four separate fill in reactions can be performed where each reaction contains one fluorescently labeled nucleotide (ddATP, ddTTP, ddGTP, or ddCTP) and unlabeled ddNTPs (see Example 2, FIGS. 7A-7D and FIGS. 9A-C). Four separate "fill in" reactions will allow detection of any nucleotide that is present at the loci of interest. For example, if analyzing a sample that contains multiple loci of interest from a single individual, and said individual is heterozygous at one or more than one loci of interest, four separate "fill in" reactions can be used to determine the nucleotides at the heterozygous loci of interest.

Also, when analyzing a sample that contains templates from multiple individuals, four separate "fill in" reactions will allow detection of nucleotides present in the sample, independent of how frequent the nucleotide is found at the locus of interest. For example,

if a sample contains DNA templates from 50 individuals, and 49 of the individuals have a thymidine at the locus of interest, and one individual has a guanine, the performance of four separate "fill in" reactions, wherein each "fill in" reaction is run in a separate lane of a gel, such as in FIGS. 9A-9C, will allow detection of the guanine. When analyzing a sample comprised of multiple DNA templates, multiple "fill in" reactions will alleviate the need to distinguish multiple nucleotides at a single site of interest by differences in mass.

In this example, multiple single nucleotide polymorphisms were analyzed. It is also possible to determine the presence or absence of mutations, including but not limited to point mutations, transitions, transversions, translocations, insertions, and deletions from multiple loci of interest. The multiple loci of interest can be from a single chromosome or from multiple chromosomes. The multiple loci of interest can be from a single gene or from multiple genes.

The sequence of multiple loci of interest that cause or predispose to a disease phenotype can be determined. For example, one could amplify one to tens to hundreds to thousands of genes implicated in cancer or any other disease. The primers can be designed so that each amplified loci of interest differs in size. After PCR, the amplified loci of interest can be combined and treated as a single sample. Alternatively, the multiple loci of interest can be amplified in one PCR reaction or the total number of loci of interest, for example 100, can be divided into samples, for example 10 loci of interest per PCR reaction, and then later pooled. As demonstrated herein, the sequence of multiple loci of interest can be determined. Thus, in one reaction, the sequence of one to ten to hundreds to thousands of genes that predispose or cause a disease phenotype can be determined.

EXAMPLE 4

The ability to determine the sequence or detect chromosomal abnormalities of a fetus using free fetal DNA in a sample from a pregnant female has been hindered by the low percentage of free fetal DNA. Increasing the percentage of free fetal DNA would enhance the detection of mutation, insertion, deletion, translocation, transversion, monosomy, trisomy, trisomy 21, trisomy 18, trisomy 13, XXY, XXX, other aneuoplodies, deletion, addition, amplification, translocation and rearrangement. The percent of fetal DNA in plasma obtained from a pregnant female was determined both in

the absence and presence of inhibitors of cell lysis. A genetic marker on the Y chromosome was used to calculate the percent of fetal DNA.

Preparation of Template DNA

The DNA template was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. The blood was aliquoted into two tubes (Fischer Scientific, 9 ml EDTA Vacuette tubes, catalog number NC9897284). Formaldehyde (25 µl/ml of blood) was added to one of the tubes. The sample in the other tube remained untreated, except for the presence of the EDTA. The tubes were spun at 1000 rpm for ten minutes. Two milliliters of the supernatant (the plasma) of each sample was transferred to a new tube and spun at 3000 rpm for ten minutes. 800 µl of each sample was used for DNA purification. DNA was isolated using the Qiagen Midi Kit for purification of DNA from blood cells (QIAmp DNA Blood Midi Kit, Catalog number 51183). DNA was eluted in 100 µl of distilled water. Two DNA templates were obtained: one from the blood sample treated with EDTA, and one from the blood sample treated with EDTA and formaldehyde.

Primer Design

Two different sets of primers were used: one primer set was specific for the Y chromosome, and thus specific for fetal DNA, and the other primer set was designed to amplify the cystic fibrosis gene, which is present on both maternal template DNA and fetal template DNA.

In this example, the first and second primers were designed so that the entire 5' and 3' sequence of each primer annealed to the template DNA. In this example, the fetus had an XY genotype, and the Y chromosome was used as a marker for the presence of fetal DNA. The following primers were designed to amplify the SRY gene on the Y chromosome.

First primer:

5' TGGCGATTAAGTCAAATTCGC 3'

Second primer:

5 CCCCCTAGTACCCTGACAATGTATT 3'

Primers designed to amplify any gene, or region of a region, or any part of any chromosome could be used to detect maternal and fetal DNA. In this example, the following primers were designed to amplify the cystic fibrosis gene:

First primer:

5' CTGTTCTGTGATATTATGTGTGGT 3'

Second primer:

5' AATTGTTGGCATTCCAGCATTG 3'

PCR Reaction

The SRY gene and the cystic fibrosis gene were amplified from the template genomic DNA using PCR (U.S. Patent Nos. 4,683,195 and 4,683,202). For increased specificity, a "hot-start" PCR was used. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by Qiagen (Catalog No. 203443). For amplification of the SRY gene, the DNA eluted from the Qiagen purification column was diluted serially 1:2. For amplification of the cystic fibrosis gene, the DNA from the Qiagen purification column was diluted 1:4, and then serially diluted 1:2. The following components were used for each PCR reaction: 8 µl of template DNA (diluted or undiluted), 1 µl of each primer (5 µM), 10 µl of HotStar Taq mix. The following PCR conditions were used:

(1) 950C for 15'

(2) 94°C for 1'

(3) 54°C for 15"

(4) 72°C for 30"

(5) Repeat steps 2-4 for 45 cycles.

(6) 10' at 72°C

Quantification of Fetal DNA

The DNA templates that were eluted from the Qiagen columns were serially diluted to the following concentrations: 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256, 1:512, 1:1024, 1:2048, and 1:4096. Amplification of the SRY gene was performed using the templates that were undiluted, 1:2, 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256, 1:512. Amplification of the cystic fibrosis gene was performed using the DNA templates that were diluted 1:4, 1:8, 1:16, 1:32, 1:64, 1:128, 1:256, 1:512, 1:1024, 1:2048, and 1:4096. The same dilution series was performed with the DNA templates that were purified from the plasma sample treated with EDTA alone and the plasma sample treated with EDTA and formaldehyde.

The results of the PCR reactions using the DNA template that was isolated from the plasma sample treated with EDTA are shown in FIG. 11A. The SRY gene was amplified from the undiluted DNA template, and also in the sample that was diluted 1:2

(FIG. 11A). The SRY gene was not amplified in the next seven serial dilutions. On the other hand, the cystic fibrosis gene was detected in the serial dilutions up to 1:256. A greater presence of the cystic fibrosis gene was expected because of the higher percentage of maternal DNA present in the plasma. The last dilution sample that provided for amplification of the gene product was assumed to have one copy of the cystic fibrosis gene or the SRY gene.

The results of the PCR reactions using the DNA template that was isolated from the plasma sample treated with formaldehyde and EDTA are shown in FIG. 11B. The SRY gene was amplified from the undiluted DNA template, and also in the sample that was diluted 1:2 (FIG. 11B). The SRY gene was not amplified in the next six dilutions. However, in the 1:256 dilution, the SRY gene was detected. It is unlikely that the amplification in the 1:256 sample represents a real signal because the prior six dilution series were all negative for amplification of SRY. Amplification of the SRY gene in this sample was likely an experimental artifact resulting from the high number of PCR cycles used. Thus, the 1:256 sample was not used in calculating the amount of fetal DNA present in the sample.

Amplification of the cystic fibrosis gene was detected in the sample that was diluted 1:16 (FIG. 11B). The presence of the formalin prevents maternal cell lysis, and thus, there is a lower percentage of maternal DNA in the sample. This is in strong contrast to the sample that was treated with only EDTA, which supported amplification up to a dilution of 1:256.

The percent of fetal DNA present in the maternal plasma was calculated using the following formula:

% fetal DNA = (amount of SRY gene/amount of cystic fibrosis gene)*2* 100.

The amount of SRY gene was represented by the highest dilution value in which the gene was amplified. Likewise, the amount of cystic fibrosis gene was represented by the highest dilution value in which it was amplified. The formula contains a multiplication factor of two (2), which is used to normalize for the fact that there is only one copy of the SRY gene (located on the Y chromosome), while there are two copies of the cystic fibrosis gene.

For the above example, the percentage of fetal DNA present in the sample that was treated with only EDTA was 1.56 % (2/256 * 2 *100). The reported percentage of fetal DNA present in the plasma is between 0.39-11.9 % (Pertl and Bianchi, *Obstetrics*

and Gynecology, Vol. 98, No. 3, 483-490 (2001). The percentage of fetal DNA present in the sample treated with formalin and EDTA was 25% (2/16 * 2 * 100). The experiment was repeated numerous times, and each time the presence of formalin increased the overall percentage of fetal DNA.

The percent fetal DNA from eighteen blood samples with and without formalin was calculated as described above with the exception that serial dilutions of 1:5 were performed. As 1:5 dilutions were performed, the last serial dilution that allowed detection of either the SRY gene or the cystic fibrosis gene may have had one copy of the gene or it may have had 4 copies of the gene. The results from the eighteen samples with and without formalin are summarized in Table V. The low range assumes that the last dilution sample had one copy of the genes and the high range assumes that the last dilution had four copies of the genes.

Table V. Mean Percentage Fetal DNA with and without formalin.

Sample	Lower Range	Upper Range
Formalin	19.47	43.69
Without Formalin	7.71	22.1

An overall increase in fetal DNA was achieved by reducing the maternal cell lysis, and thus, reducing the amount of maternal DNA present in the sample. In this example, formaldehyde was used to prevent lysis of the cells, however any agent that prevents the lysis of cells or increases the structural integrity of the cells can be used. Two or more than two cell lysis inhibitors can be used. The increase in fetal DNA in the maternal plasma allows the sequence of the fetal DNA to be determined, and provides for the rapid detection of abnormal DNA sequences or chromosomal abnormalities including but not limited to point mutation, reading frame shift, transition, transversion, addition, insertion, deletion, addition-deletion, frame-shift, missense, reverse mutation, and microsatellite alteration, trisomy, monosomy, other aneuploidies, amplification, rearrangement, translocation, transversion, deletion, addition, amplification, fragment, translocation, and rearrangement.

EXAMPLE 5

A DNA template from an individual with a genotype of trisomy 21 was analyzed. Three loci of interest were analyzed on chromosome 13 and two loci of interest were analyzed on chromosome 21.

Preparation of Template DNA

The template DNA was prepared from a 5 ml sample of blood obtained by venipuncture from a human volunteer with informed consent. The human volunteer had previously been genotyped to have an additional chromosome 21 (trisomy 21). Template DNA was isolated using QIAamp DNA Blood Midi Kit supplied by QIAGEN (Catalog number 51183).

Primer Design

The following five single nucleotide polymorphisms were analyzed: SNP TSC 0115603 located on chromosome 21; SNP TSC 03209610 located on chromosome 21; SNP TSC 0198557 located on chromosome 13; and SNP TSC 0200347 located on chromosome 13. The DNA template from another individual was used as an internal control. The SNP TSC 0200347, which was previously identified as being homozygous for guanine, was used as the internal control. The SNP Consortium Ltd database can be accessed at http://snp.cshl.org/, website address effective as of April 1, 2002.

SNP TSC 0115603 was amplified using the following primers:

First Primer:

5' GTGCACTTACGTGAATTCAGATGAACGTGATGTAGTAG 3'

Second Primer:

5' TCCTCGTACTCAACGGCTTTCTCTGAAT 3'

The first primer was biotinylated at the 5' end, and contained the restriction enzyme recognition site for EcoR I. The second primer contained the restriction enzyme recognition site for the restriction enzyme BceA I.

SNP TSC 0309610 was amplified using the following primers:

First primer:

5' TCCGGAACACTAGAATTCTTATTTACATACACACTTGT 3'

Second primer:

5' CGAATAAGGTAGACGGCAACAATGAGAA 3'

The first primer contained a biotin group at the 5' end, and a restriction enzyme recognition site for the restriction enzyme EcoR I. The second primer contained the restriction enzyme recognition site for BceA I.

Submitted SNP (ss) 813773 (accession number assigned by the NCBI Submitted SNP (ss) Database) was amplified with the following primers:

First primer:

5' CGGTAAATCGGAGAATTCAGAGGATTTAGAGGAGCTAA 3'

Second primer:

5' CTCACGTTCGTTACGGCCATTGTGATAGC 3'

The first primer contains a biotin group at the 5' end, and a recognition site for the restriction enzyme EcoR I. The second primer contained the restriction enzyme recognition site for BceA I.

SNP TSC 0198557 was amplified with the following primers:

First primer:

5' GGGGAAACAGTAGAATTCCATATGGACAGAGCTGTACT 3'

Second primer:

5' TGAAGCTGTCGGACGGCCTTTGCCCTCTC 3'

The first primer contains a biotin group at the 5' end, and a recognition site for the restriction enzyme EcoR I. The second primer contained the restriction enzyme recognition site for BceA I.

SNP TSC 0197279 was amplified with the following primers:

First primer:

5' ATGGGCAGTTATGAATTCACTACTCCCTGTAGCTTGTT 3'

Second primer:

5' TGATTGGCGCGAACGGCACTCAGAGAAGA 3'

The first primer contained a biotin group at the 5' end, and a recognition site for the restriction enzyme for EcoR I. The second primer contained the recognition site for the restriction enzyme BceA I.

SNP TSC 0200347 was amplified with the following primers:

First primer:

5' CTCAAGGGGACCGAATTCGCTGGGGTCTTCTGTGGGTC 3'

Second primer:

5' TAGGGCGGCGTGACGGCCAGCCAGTGGT 3'

The first primer contained a biotin group at the 5' end, and the recognition site for the restriction enzyme EcoR I. The second primer contained the restriction enzyme recognition site for BceA I.

PCR Reaction

All five loci of interest were amplified from the template genomic DNA using PCR (U.S. Patent Nos. 4,683,195 and 4,683,202). For increased specificity, a "hot-start" PCR was used. PCR reactions were performed using the HotStarTaq Master Mix Kit supplied by QIAGEN (catalog number 203443). The amount of template DNA and primer per reaction can be optimized for each locus of interest; in this example, 40 ng of template human genomic DNA and 5 μM of each primer were used. Thirty-eight cycles of PCR were performed. The following PCR conditions were used for SNP TSC 0115603, SNP TSC 0309610, and SNP TSC 02003437:

(1) 95°C for 15 minutes and 15 seconds;
(2) 42°C for 30 seconds;
(3) 95°C for 30 seconds;
(4) 60°C for 30 seconds;
(5) 95°C for 30 seconds;
(6) 69°C for 30 seconds;
(7) 95°C for 30 seconds;
(8) Repeat steps 6 and 7 thirty nine (37) times;
(9) 72°C for 5 minutes.

The following PCR conditions were used for SNP ss813773, SNP TSC 0198557, and SNP TSC 0197279:

(1) 95°C for 15 minutes and 15 seconds;
(2) 37°C for 30 seconds;
(3) 95°C for 30 seconds;
(4) 57°C for 30 seconds;
(5) 95°C for 30 seconds;
(6) 64°C for 30 seconds;
(7) 95°C for 30 seconds;
(8) Repeat steps 6 and 7 thirty nine (37) times; and
(9) 72°C for 5 minutes.

In the first cycle of each PCR, the annealing temperature was about the melting temperature of the 3' annealing region of the second primer. The annealing temperature in the second cycle of PCR was about the melting temperature of the 3' region, which anneals to the template DNA, of the first primer. The annealing temperature in the third cycle of PCR was about the melting temperature of the entire sequence of the second primer. Escalating the annealing temperature from TM1 to TM2 to TM3 in the first three cycles of PCR greatly improves specificity. These annealing temperatures are representative, and the skilled artisan will understand the annealing temperatures for each cycle are dependent on the specific primers used. The temperatures and times for denaturing, annealing, and extension, can be optimized by trying various settings and using the parameters that yield the best results.

Purification of Fragment of Interest

PCR products were separated from the components of the PCR reaction using Qiagen's MinElute PCR Purification Kit following manufacturer's instructions (Catalog number 28006). The PCR products were eluted in 20 µl of distilled water. For each amplified SNP, one microliter of PCR product, 1 µl of amplified internal control DNA (SNP TSC 0200347), and 8 µl of distilled water were mixed. Five microliters of each sample was placed into two separate reaction wells of a Pierce StreptaWell Microtiter plate (catalog number 15501). The first primers contained a 5' biotin tag so the PCR products bound to the Streptavidin coated wells while the genomic template DNA did not. The streptavidin binding reaction was performed using a Thermomixer (Eppendorf) at 150 rpm for 1 hour at 45°C. Each well was aspirated to remove unbound material, and washed three times with 1X PBS, with gentle mixing (Kandpal et al., Nucl. Acids Res. 18:1789-1795 (1990); Kaneoka et al., Biotechniques 10:30-34 (1991); Green et al., Nucl. Acids Res. 18:6163-6164 (1990)).

Restriction Enzyme Digestion of Isolated Fragments

The purified PCR products were digested with the restriction enzyme that bound the recognition site that was incorporated into the PCR products from the second primer. The purified PCR products were digested with the restriction enzyme BceA I (New England Biolabs, catalog number R0623S). The digests were performed in the wells of the microtiter plate following the instructions supplied with the restriction enzyme. After

digestion with the appropriate restriction enzyme, the wells were washed three times with PBS to remove the cleaved fragments.

Incorporation of Labeled Nucleotide

The restriction enzyme digest described above yielded a DNA fragment with a 5' overhang, which contained the SNP and a 3' recessed end. The 5' overhang functioned as a template allowing incorporation of a nucleotide or nucleotides in the presence of a DNA polymerase.

For each SNP, two fill in reactions were performed; each reaction contained a different fluorescently labeled ddATP (ddATP, ddATP, ddATP, or ddATP, depending on the reported nucleotides to exist at a particular SNP). For example, the nucleotides adenine and thymidine have been reported at SNP TSC 0115603. Therefore, the digested PCR product for SNP TSC 0115603 was mixed with either fluorescently labeled ddATP or fluorescently labeled ddATP. Each reaction contained fluorescently labeled ddATP for the internal control. The following components were added to each fill in reaction: 2 µl of a ROX-conjugated ddATP (depending on the nucleotides reported for each SNP), 2 µl of ROX-conjugated ddATP (internal control), 2.5 µl of 10X sequenase buffer, 2 µl of Sequenase, and water as needed for a 25 µl reaction. All of the fill in reactions were performed at 45°C for 45 min. However, shorter time periods of incorporation can be used. Non-fluorescently labeled ddNTPs were purchased from Fermentas Inc. (Hanover, MD). The ROX-conjugated ddNTPs were obtained from Perkin Elmer. In the presence of fluorescently labeled ddNTPs, the 3' recessed end was extended by one base, which corresponds to the SNP or locus of interest.

After labeling, each Streptawell was rinsed with 1X PBS (100 µl) three times. The "filled in" DNA fragments were then released from the Streptawells by digestion with the restriction enzyme EcoR I following manufacturer's recommendations. Digestion was performed for 1 hour at 37°C with shaking at 120 rpm.

Detection of the Locus of Interest

After release from the streptavidin matrix, 3 µl of the 10 µl sample was loaded in a 48 well membrane tray (The Gel Company, catalog number TAM48-01). The sample in the tray was absorbed with a 48 Flow Membrane Comb (The Gel Company, catalog

number AM48), and inserted into a 36 cm 5% acrylamide (urea) gel (BioWhittaker Molecular Applications, Long Ranger Run Gel Packs, catalog number 50691).

The sample was electrophoresed into the gel at 3000 volts for 3 min. The membrane comb was removed, and the gel was run for 3 hours on an ABI 377 Automated Sequencing Machine. The incorporated labeled nucleotide was detected by fluorescence.

As seen in FIG. 12, SNP TSC 0115603 was "filled in" with labeled ddTTP (lane 1) and in a separate reaction with labeled ddATP (lane 3). The calculated ratio between the nucleotides, using the raw data, was 66:34, which is consistent with the theoretical ratio of 66:33 for a SNP on chromosome 21 in an individual with trisomy 21. Both the ddTTP and ddATP were labeled with the same fluorescent dye to minimize variability in incorporation efficiencies of the dyes. However, nucleotides with different fluorescent labels or any detectable label can be used. It is preferable to calculate the coefficients of incorporation when different labels are used.

Each fill in reaction was performed in a separate well so it was possible that there could be variability in DNA binding between the wells of the microtiter plate. To account for the potential variability of DNA binding to the streptavidin-coated plates, an internal control was used. The internal control (SNP TSC 0200347), which is homozygous for guanine, was added to the sample prior to splitting the sample into two separate wells, and thus, an equal amount of the internal control should be present in each well. The amount of incorporated ddGTP can be fixed between the two reactions. If the amount of DNA in each well is equal, the amount of incorporated ddGTP should be equal because the reaction is performed under saturating conditions, with saturating conditions being defined as conditions that support incorporation of a nucleotide at each template molecule. Using the internal control, the ratio of incorporated ddATP to ddTTP was 63.4:36.6. This ratio was very similar to the ratio obtained with the raw data, indicating that there are minor differences in the two fill in reactions for a particular SNP.

Table VI. Allele Frequencies at Multiple SNPs on DNA Template from Individual with Trisomy 21

SNP	Allele	Peak Area	Allele Ratio	Internal Control	Normalized Peak Area	Allele Ratio (%)
TSC 0115603	A	5599	66	723	5599	63.4
	T	2951	34	661	3227 ((723/661)*2951)	36.6

TSC 0309610	T C	4126 2342	64 36	1424 1631	4126 2045 ((1424/1631)*2342)	66.8 33.2
ss813773	A C	4199 4870	46 54	808 647	4199 6082 ((808/647)*4870)	41 59
TSC 0198557	T C	3385 2741	55 45	719 559	3385 3525 719/559 *2741)	49 51
TSC 0197279	T C	8085 7202	53 47	2752 2520	8085 7865 (2752/2520 *7202	50.7 49.3

SNP TSC 0309610 was filled in with ddTTP (lane 3) or ddCTP (lane 4) (FIG. 12). The calculated ratio for the nucleotides, using the raw data, was 64:36. Both ddTTP and ddCTP were labeled with the same fluorescent dye. After normalization to the internal control, as discussed above, the calculated allele ratio of ddTTP to ddCTP was 66.8:33.2 (Table VI). Again, the both the calculated ratio from the raw data and the calculated ratio using the internal control are very similar to the theoretical ratio of 66.6:33.4 for a SNP on chromosome 21 in an individual with trisomy.

To demonstrate that the 66:33 ratios for nucleotides at heterozygous SNPS represented loci on chromosomes present in three copies, SNPs on chromosome 13 were analyzed. The individual from whom the blood sample was obtained had previously been genotyped with one maternal chromosome 13, and one paternal chromosome 13.

Submitted SNP (ss) 813773 was filled in with ddATP (lane 5) or ddCTP (lane 6) (FIG. 12). The calculated ratio for the nucleotides at this heterozygous SNP, using the raw data, was 46:54. This ratio is within 10% of the expected ratio of 50:50. Importantly, the ratio does not approach the 66:33 ratio expected when there is an additional copy of a chromosome.

After normalization to the internal control, the calculated ratio was 41:59. Contrary to the expected result, normalization to the internal control increased the discrepancy between the calculated ratio and the theoretical ratio. This result may represent experimental error that occurred in aliquoting the DNA samples.

Also, it is possible that the restriction enzyme used to generate the overhang, which was used as a template for the "fill-in" reaction, preferentially cut one DNA template over the other DNA template. The two templates differ, with respect to the nucleotide at the SNP site, and this may influence the cutting. The primers can be designed such that the nucleotides adjacent to the cut site are the same, independent of the nucleotide at the SNP site (discussed further in the section entitled "Primer Design").

SNP TSC 0198557, which is on chromosome 13, was filled in with ddTTP (lane 7) in one reaction and ddCTP (lane 8) in another (FIG. 12). The calculated ratio for the nucleotides at this SNP, using the raw data, was 55:45. After normalization to the internal control, the calculated allele ratio of T:C was 49:51. The normalized ratio was closer to the theoretical ratio of 50:50 for an individual with two copies of chromosome 13.

SNP TSC 0197279, which is on chromosome 13, was filled in with ddTTP (lane 9) in one reaction and ddCTP (lane 10) in another (FIG. 12). The calculated ratio for the nucleotides at this SNP, using the raw data, was 53:47.. After normalization to the internal control, the calculated allele ratio of T:C was 50.7:49.3. This is consistent with the theoretical ratio of 50:50 for an individual with only two copies of chromosome 13.

The ratio for the nucleotides at two of the analyzed SNPs on chromosome 13 was approximately 50:50. One SNP, ss813773, showed a ratio of 46:54, and when normalized to the internal control, the ratio was 41:59. These ratios deviate from the expected 50:50, but at the same time, the ratios are not indicative of an extra chromosome, which is indicated with a ratio of 66:33. While the data from this particular SNP is inconclusive, it does not represent a false positive. No conclusion could be drawn on the data from this SNP. However, the other two SNPs provided data that indicated a normal number of chromosomes. It is preferable to analyze multiple SNPs on a chromosome including but not limited to 1-5, 5-10, 10-50, 50-100, 100-200, 200-300, 300-400, 400-500, 500-600, 600-700, 700-800, 800-900, 900-1000, 1000-2000, 2000-3000, and greater than 3000. Preferably, the average of the ratios for a particular chromosome will be used to determine the presence or absence of a chromosomal abnormality. However, it is still possible to analyze one locus of interest. In the event that inconclusive data is obtained, another locus of interest can be analyzed.

The individual from whom the DNA template was obtained had previously been genotyped with trisomy 21, and the allele frequencies at SNPs on chromosome 21 indicate the presence of an additional chromosome 21. The additional chromosome contributes an additional nucleotide for each SNP, and thus alters the traditional 50:50 ratio at a heterozygous SNP. These results are consistent for multiple SNPs, and are specific for those found on chromosome 21. The allele frequencies for SNPs on chromosome 13 gave the expected ratios of approximately 50:50. These results demonstrate that this method of SNP detection can be used to detect chromosomal